THESE DE DOCTORAT DE

L'ECOLE NATIONALE SUPERIEURE
DE TECHNIQUES AVANCEES BRETAGNE
COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies
de l'Information et de la Communication
Spécialité : Signal, Image, Vision

Par

Clément J. Roussel
**Stochastic differential equations for the electromagnetic field
scattered by the sea surface :** applications to remote sensing

Thèse présentée et soutenue à Brest, le 4 septembre 2019
Unité de recherche : Lab-STICC UMR CNRS6285

Rapporteurs avant soutenance :
Josselin GARNIER, Professeur, Ecole Polytechnique Palaiseau
Gabriel SORIANO, Maître de Conférences, HDR, Institut Fresnel Marseille

Composition du Jury :
Président : Thierry CHONAVEL, Professeur, IMT-Atlantique Brest
Examinateurs :
Josselin GARNIER,   Professeur, Ecole Polytechnique Palaiseau
Gabriel SORIANO,   Maître de Conférences, HDR, Université Aix-Marseille
Thierry CHONAVEL,   Professeur, IMT-Atlantique Brest
Marie DU ROY DE CHAUMARAY, Maître de Conférences, ENSAI Rennes
Charles-Antoine GUERIN,   Professeur des Universités, Université de Toulon

Directeur de thèse :
Alexandre BAUSSARD,   Professeur des Universités, Université de Technologie de Troyes
Co-encadrant :
Arnaud COATANHAY,   Enseignant-Chercheur, ENSTA Bretagne Brest
Invité
Philippe POULIGUEN,   Responsable Domaine Ondes Acoustiques et Radioélectriques, HDR,
                                  Agence de l'Innovation de Défense

Remerciements

Plus qu'un choix, faire une thèse s'est imposé à moi comme une véritable nécessité. Celle d'aller plus loin, de se faire mal à l'esprit, de comprendre davantage les ficelles de la science, et finalement de contribuer un tant soit peu à celle-ci. Alexandre Baussard et Arnaud Coatanhay m'ont donné l'opportunité de vivre cette expérience, ils m'ont fait confiance et m'ont accepté dans une situation d'échec personnel où faire de la recherche ne me semblait tout simplement plus envisageable. À ce titre, ils méritent mes premiers remerciements. Alexandre, nous avons mis du temps à nous accorder (mais nous avons réussi!) et parfois il y a eu des étincelles entre nous. Au fond, je suis convaincu que c'était pour le mieux, c'est dans cette adversité qu'on a pû être productifs. Je te remercie en particulier de m'avoir transmis ta rigueur et ton souci du détail. J'avais un sacré chemin à faire, et je crois m'être sensiblement amélioré grâce à toi! Arnaud, je n'ai pas épuisé toutes les questions que j'avais à te poser et j'imagine que je n'ai d'autre choix que de renoncer maintenant. Merci d'avoir répondu patiemment aux plus bêtes et aux plus excentriques de mes interrogations. Tu m'as transmis une manière d'aborder la science bien éloignée de mon fantasme initial d'un livre qui contiendrait tout ce qu'il y a à connaître. Merci également au soutien administratif assuré par Annick et Michèle! Comment aurais-je fais sans vous?

Je remercie deux protagonistes sans lesquels rien ne se serait passé: Amandine Nicolle, qui m'a toujours soutenu quand j'étais élève ingénieur et a encore soutenu ma candidature pour cette thèse, et mon ami à la langue bien (heureusement!) pendue Antoine Parent (et merci aussi à Florian Schattner pour qu'il n'y ait pas de jaloux)!

Merci également à la DGA pour avoir soutenu financièrement le projet, et à mon tuteur Philippe Pouliguen.

Faire une thèse ne nécessite pas seulement que les bonnes conditions "matérielles" soient réunies. Faire une thèse, c'est s'engouffrer dans un tunnel long de trois ans et cheminer laborieusement. C'est faire de gros sacrifices et se retrouver parfois terriblement seul, prisonnier de ses propres pensées, face à soi-même. Traverser ça suppose de pouvoir respirer, se ressourcer. Merci donc à tous mes amis. Je pense aux brestois qui m'ont aidé à supporter la bruine, en particulier Chris pour ses petits plats cuisinés sous vide que j'ai eu plaisir à déguster devant des épisodes d'Archer, un bon verre de rouge à la main et un rhum jamaicain en dessert, David pour son sourire et ses yeux étincelants devant lequels il est impossible de rester déprimé, et Carl pour sa discussion et sa géométrie algébrique imbitable qui a le bon goût de me remettre illico presto à ma place! Merci à bien d'autres également qui se reconnaîtront et me pardonnerons je l'espère de ne pas tous les citer. Merci à mes amis de longue date, à Maxime Thivolle pour les nombreuses corrections qu'il a apporté à ma thèse en ayant l'invraisemblable patience de la relire, à tous les membres du "Clan" et en particulier Memert, mon plus ancien acolyte, meilleur ami d'enfance, tueur de grenouilles, partenaire de badminton et maintenant fournisseur officiel de vinasse gouleyante et trébuchante. Merci à mon meilleur ami d'esprit, mon *alter ego*, Simon Quemin (PhD), qui en mettant la barre très haut et en ayant toujours une longueur d'avance m'a poussé à me dépasser. On a bien d'autres discussions Bitboliennes et Stiegleriennes à poursuivre, tout en escaladant les montagnes d'une main et en disputant une partie de squash de l'autre. Pendant tout ce temps, il va sans dire qu'on est alimenté par un casque à boisson avec de l'IPA d'un côté et du Pu'er de l'autre.

"Car il se trouve en effet des gens pour dire que si l'on met une casserole d'eau sur le feu, il n'est que très fortement probable que l'eau bouille et ne gèle pas, et qu'à strictement parler ce que nous tenons pour impossible n'est qu'improbable. Quelle différence cela fait-il dans leur vie? N'est-ce pas seulement que, au sujet de certaines choses, ils parlent plus que nous autres?"

Ludwig Wittgenstein, De la certitude

"For there are people who say that it is merely extremely probable that water over a fire will boil and not freeze, and that therefore strictly speaking what we consider impossible is only improbable. What difference does this make in their lives? Isn't it just that they talk rather more about certain things than the rest of us?"

Ludwig Wittgenstein, On certainty

Approche stochastique pour la diffusion électromagnétique par la surface de la mer : application à la télédétection

Résumé étendu

1   Introduction

La télédétection en milieu maritime (au dessus de la surface) a des applications diverses dont l'intérêt est indéniable : surveillance de l'espace maritime (lutte contre les activités clandestines, protection de l'environnement), océanographie, suivi des glaces de mer etc. Parmi les systèmes d'observations possibles, le radar (Radio Detection and Ranging) est particulièrement utilisé pour faire de la détection, ou de l'imagerie. Les avantages du radar actif sur un capteur optique passif par exemple, sont sa capacité à fonctionner de jour comme de nuit, et son insensibilité relative aux conditions météorologiques (les ondes radar traversent la couverture nuageuse). Plusieurs plateformes peuvent être utilisées : côtier (sur une falaise), bateau, aéroporté, satellitaire etc.

La caractéristique majeure de la télédétection en milieu maritime est la présence du signal rétrodiffusé par la mer, qu'on appelle clutter de mer (reflectivité complexe de la mer). Le clutter de mer est un signal utile lorsque la finalité est l'étude de la mer, mais un bruit pour la détection ou l'imagerie de cibles, en raison de son caractère complexe et imprévisible (voir figure 1).
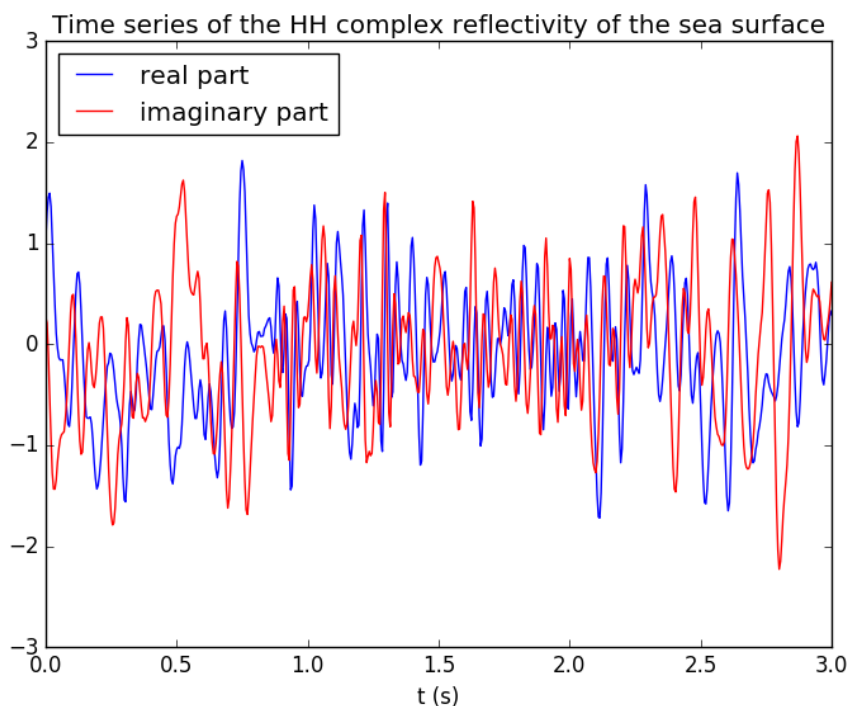
FIGURE 1 – Séries temporelles de la réfléctivité complexe de la mer (clutter de mer) enregistrées par le radar en bande C de l'Ifremer, en polarisation HH.

Une compréhension et description profondes de la nature du clutter de mer sont souhaitables quelle que soit la finalité recherchée. De nombreux travaux ont donc été menés depuis plusieurs décennies, et certains modèles ont gagné leurs lettres de noblesse pour la finalité détection : les modèles statistiques. Ces types de modèles décrivent le clutter de mer par une distribution de probabilité. Si on note Ψ_t la réfléctivité complexe de la mer, et $z_t = |\Psi_t|^2$ l'intensité, la K distribution [17] a été utilisée pour modéliser l'intensité. Elle se déduit théoriquement en considérant que la surface de la mer est constituée d'une population de

diffuseurs discrets, telle que le nombre de diffuseurs suit une loi binomiale négative et telle que la phase de chaque diffuseur est uniformément répartie sur $[0, 2\pi[$, et en faisant tendre le nombre de diffuseurs vers l'infini. Ce modèle de diffuseurs discrets est une marche aléatoire [5].

La force de cette approche statistique directe est qu'elle permet par exemple de rapidement obtenir des algorithmes de détection. En revanche, c'est une approche statique dans le sens où si elle modélise la distribution de probabilité de z_t quel que soit t fixé, elle n'apporte aucune connaissance sur la corrélation entre z_{t_1} et z_{t_2} pour $t_1 \neq t_2$, ce qui mène à supposer de manière non-appropriée que différents échantillons du clutter de mer sont indépendants. Nous pensons qu'un modèle dynamique serait préférable pour décrire le clutter de mer, puisqu'il est lui-même fondamentalement dynamique.

Un tel modèle a récemment été développé par T. R. Field [4]. Il représente le clutter de mer par des processus aléatoires décrits par des équations différentielles stochastiques (EDS). Le modèle de Field est une extension dynamique de la K distribution/marche aléatoire puisqu'il affirme toujours que z_t est K-distribué pour tout t, mais qu'il prend en compte la dépendance temporelle du clutter de mer grâce aux EDS. Le modèle de Field hérite donc des avantages de la K distribution, mais en repousse les limites en proposant une solution à son problème de staticité.

Nos travaux de recherche se placent dans le cadre général de la télédétection en milieu maritime. Nous proposons d'étudier et de développer le modèle de Field afin de le transformer en cadre et outil utilisables dans le traitement de séries temporelles du clutter de mer. Les applications principalement visées sont la détection et l'imagerie Radar à Synthèse d'Ouverture (RSO).

Dans la section 2, nous présentons quelques élements de calcul stochastique, en particulier les équations différentielles stochastiques. Dans la section 3, nous introduisons le modèle de Field pour le clutter de mer en montrant son lien avec la marche aléatoire. Dans la section 4, nous résolvons des équations de Fokker-Planck pour obtenir des expressions analytiques de probabilités de transition, et proposons de les utiliser comme outil de synchronisation de données. Dans la section 5, nous estimons les paramètres du modèle de Field pour le clutter de mer (par maximum de vraisemblance, et moments) en faisant une hypothèse d'observabilité de la surface équivalente radar de la mer (SER). Dans la section 6, nous adaptons les EDS du modèle de Field pour y inclure une cible, et estimons les paramètres de la cible par maximum de vraisemblance. Dans la section 7, nous travaillons sur la non-observabilité de la SER de la mer en pratique (en écho à la section 5). Nous montrons qu'il est possible de l'estimer à partir de données observables, puis reprenons l'estimation des paramètres du clutter en proposons une séquence d'estimateurs applicable à des données réelles. La section 8 conclut.

N.B : L'organisation de ce résumé suit celle de la thèse. La section 1, que nous clôturont ici, correspond à l'introduction et au chapitre 1. Les sections 2 à 7 correspondent respectivement aux chapitres 2 à 7, et la section 8 correspond à la conclusion.

2   Calcul stochastique

Dans cette section, nous présentons quelques notions de calcul stochastique nécessaires pour comprendre le modèle de Field et nos travaux. Soit X_t un processus aléatoire (famille

de variables aléatoires indexée par $t \in \mathbb{R}^+$). Une équation différentielle stochastique (EDS) pour X_t est de la forme :

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \tag{1}$$

où W_t est un mouvement brownien. On appelle μ la *dérive* et σ la *volatilité*. La définition rigoureuse des EDS fait appel au calcul d'Itô (voir [11] ou [8] par exemple). Cependant, on peut avoir une bonne intuition de cette notion en sachant résoudre une EDS par le schéma d'Euler-Maruyama.

Soit $[0, T]$ un intervalle et $t_0 = 0 < t_1 < \cdots < t_N = T$ une subdivision de $[0, T]$. Le schéma d'Euler-Maruyama s'écrit :

$$X_{t_i} = X_{t_{i-1}} + \mu(X_{t_{i-1}})(t_i - t_{i-1}) + \sigma(X_{t_{i-1}})(W_{t_i} - W_{t_{i-1}}). \tag{2}$$

Par définition du mouvement brownien, $W_{t_i} - W_{t_{i-1}}$ est une variable aléatoire gaussienne de variance $\Delta t = t_i - t_{i-1}$. On voit que sur un incrément de temps Δt, la dérive correspond à un incrément déterministe, comme une équation différentielle ordinaire, tandis que la volatilité correspond à un incrément aléatoire. Si $X_{t_{i-1}}$ est fixé, le schéma d'Euler-Maruyama implique que :

$$X_{t_i} - X_{t_{i-1}} \sim \mathcal{N}(\Delta t \mu(X_{t_{i-1}}), \Delta t \sigma(X_{t_{i-1}})^2), \tag{3}$$

i.e. que l'incrément $X_{t_i} - X_{t_{i-1}}$ est une variable aléatoire gaussienne centrée en $\Delta t \mu(X_{t_{i-1}})$ et de variance $\Delta t \sigma(X_{t_{i-1}})^2$. La probabilité de transition du processus X_t entre t_{i-1} et t_i est donc approximativement :

$$p(X_{t_i} = x \mid X_{t_{i-1}} = y) \approx \frac{1}{\sqrt{2\pi \Delta t \sigma(y)^2}} \exp\left(-\frac{1}{2} \frac{(x - (y + \mu(y)\Delta t))^2}{\Delta t \sigma(y)^2}\right). \tag{4}$$

Cette approximation n'est valable que pour $t_i - t_{i-1}$ suffisament petit.

Le schéma d'Euler-Maruyama peut donc être utilisé à deux fins. La première est de résoudre numériquement une EDS en simulant de multiples trajectoires. Pour ça, il faut répéter l'équation (2) de t_0 à t_N pour simuler une trajectoire, puis réitérer pour générer d'autres trajectoires. Celles-ci seront différentes car les incréments browniens $W_{t_i} - W_{t_{i-1}}$ changeront d'une trajectoire à l'autre. Comme le montre l'équation (4), le schéma d'Euler-Maruyama peut aussi être utilisé pour approximer les probabilités de transition de X_t.

On peut montrer que les probabilités de transitions exactes de X_t sont solutions de l'équation de Fokker-Planck [13] :

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}[\mu(x)p] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[\sigma^2(x)p], \tag{5}$$

où $p = p(x, t)$ dépend de sa variable de distribution, du temps, et implicitement de la condition initiale. L'équation de Fokker-Planck est une équation aux dérivées partielles. Pour obtenir par exemple $p(X_{t_i} = x \mid X_{t_{i-1}} = y)$, il faut imposer une distribution de Dirac δ_y à $t = 0$ (condition initial déterministe), calculer la solution, puis l'évaluer au temps $t_i - t_{i-1}$.

A partir de l'équation de Fokker-Planck, il est également possible d'obtenir la distribution stationnaire/asymptotique en imposant $\frac{\partial p}{\partial t} = 0$ dans l'équation (5) et en la résolvant. La solution stationnaire est alors valide pour tout t et se dénote $p(X_t = x)$.

La notion de probabilité de transition est essentielle pour comprendre les sections 4 à 7.

3 Modèle de Field

Le modèle de Field se construit sur la marche aléatoire. On suppose qu'au temps t, il y a N_t diffuseurs qui contribuent à la réfléctivité complexe, où N_t suit un modèle de population Birth-Death-Immigration d'espérance $\bar{N}$. La réfléctivité complexe au temps t est donc (marche aléatoire) :

$$\Psi_t^{(\bar{N})} = \sum_{n=1}^{N_t} \frac{a}{\bar{N}^{1/2}} e^{i\phi_t^{(n)}} = \left(\frac{N_t}{\bar{N}}\right)^{1/2} \sum_{n=1}^{N_t} \frac{a}{N_t^{1/2}} e^{i\phi_t^{(n)}}. \tag{6}$$

Dans l'équation (6), l'amplitude a des diffuseurs est supposée constante. Les phases $\phi_t^{(n)}$ des diffuseurs sont indépendantes pour différents n, et elles suivent les EDS :

$$\begin{cases} \mathrm{d}\phi_t^{(n)} = \mathcal{B}^{1/2}\mathrm{d}W_t^{(n)} \\ \phi_0^{(n)} = \Delta^{(n)}, \end{cases} \tag{7}$$

où $\forall n$, $\Delta^{(n)}$ est uniformément distribuée sur $[0, 2\pi[$ et où $\mathcal{B}$ est une constante positive. En prenant $a = 1$ (normalisation) et dans la limite d'un nombre de diffuseurs infini ($\bar{N} \to +\infty$), on obtient :

$$\Psi_t = \lim_{\bar{N}\to+\infty} \left(\frac{N_t}{\bar{N}}\right)^{1/2} \lim_{\bar{N}\to+\infty} \sum_{n=1}^{N_t} \frac{1}{N_t^{1/2}} e^{i\phi_t^{(n)}}. \tag{8}$$

Dans le modèle de Field, le clutter de mer est donc représenté par le produit :

$$\Psi_t = x_t^{1/2}\gamma_t, \tag{9}$$

où $x_t = \lim_{\bar{N}\to+\infty} \frac{N_t}{\bar{N}}$ est la Surface Equivalente Radar (SER) de la surface de la mer, et où $\gamma_t = \gamma_t^{(R)} + i\gamma_t^{(I)} = \lim_{\bar{N}\to+\infty} \sum_{n=1}^{N_t} \frac{1}{N_t^{1/2}} e^{i\phi_t^{(n)}}$ est le speckle. La SER est proportionnelle au nombre de diffuseurs contribuant au clutter et évolue lentement (on parle parfois de texture). Le speckle (parfois appelé chatoiement) est le clutter en l'absence de fluctuations de population, et correspond à la décohérence de phase des diffuseurs liée à leurs mouvements. Il évolue beaucoup plus vite.

Dans le modèle de Field, x_t et γ_t sont solutions des EDS [4] :

$$\begin{cases} \mathrm{d}x_t = \mathcal{A}(1 - x_t)\mathrm{d}t + \left(2\frac{\mathcal{A}}{\alpha}x_t\right)^{\frac{1}{2}} \mathrm{d}W_t^{(x)} \\ \mathrm{d}\gamma_t^{(R)} = -\frac{1}{2}\mathcal{B}\gamma_t^{(R)}\mathrm{d}t + \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\mathrm{d}W_t^{(R)} \\ \mathrm{d}\gamma_t^{(I)} = -\frac{1}{2}\mathcal{B}\gamma_t^{(I)}\mathrm{d}t + \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\mathrm{d}W_t^{(I)}, \end{cases} \tag{10}$$

où $W_t^{(x)}$, $W_t^{(R)}$ et $W_t^{(I)}$ sont des mouvements browniens indépendants. Par rapport à [4], nous avons préféré écrire deux EDS pour les parties réelle et imaginaire du speckle, $\gamma_t^{(R)}$ et $\gamma_t^{(I)}$, plutôt qu'une EDS complexe. Ces deux processus aléatoires sont solutions d'EDS qui ont la même forme, mais qui sont supposées indépendantes par indépendance des mouvements browniens.

Le modèle de Field est paramétrisé par trois paramètres constants : $\mathcal{A}$, $\mathcal{B}$ et α. $\mathcal{A}$ (Hz) est l'inverse du temps de corrélation de x_t, tandis que $\mathcal{B}$ est l'inverse du temps de corrélation de γ_t. On sait empiriquement que x_t évolue beaucoup plus lentement que γ_t, ce qui implique qu'en pratique :

$$\mathcal{A} \ll \mathcal{B}. \tag{11}$$

α, quant à lui, est l'inverse de la variance stationnaire de x_t. On sait empiriquement (voir [17]) que le temps de corrélation de x_t est de l'ordre de la seconde et celui de γ_t de l'ordre de la dizaine de milliseconde. On peut donc considérer que des ordres de grandeur représentatifs sont $\mathcal{A} = 1$ Hz et $\mathcal{B} = 100$ Hz.

Il est important de noter que pour caractériser la réflectivité complexe Ψ_t de la surface de la mer, trois processus sont nécessaires, x_t, $\gamma_t^{(R)}$ et $\gamma_t^{(I)}$, malgré que Ψ_t soit bidimensionnel.

4   Inférences sur le clutter de mer

Dans cette partie, nous présentons une contribution de notre recherche ciblant les radars à synthèse d'ouverture (RSO) [2], [10], [9]. Les RSO, aéroportés ou satellitaires, sont utilisés pour créer des images radars haute résolution de la terre ou de la mer. Pour ce faire, le RSO survole une zone et exploite le mouvement du porteur pour créer une antenne synthétique de grande dimension qui améliore la résolution en azimuth, grâce à un processus d'intégration de différents pulses. Cette technique fonctionne grâce au fait que chaque point au sol qui est illuminé par le radar, l'est depuis une série de positions successives selon des points de vue différents. Si le RSO permet d'obtenir des images de bonne qualité sur les zones terrestres, ce n'est pas toujours le cas pour les zones maritimes du fait du mouvement de la mer pendant le processus d'intégration, ce qui crée un flou [6], [7]. Nous n'avons pas travaillé strictement sur le RSO, mais plutôt considéré la situation plus générale d'un capteur mobile qui effectue des mesures du clutter de mer à des temps et positions différentes (voir figure 2).
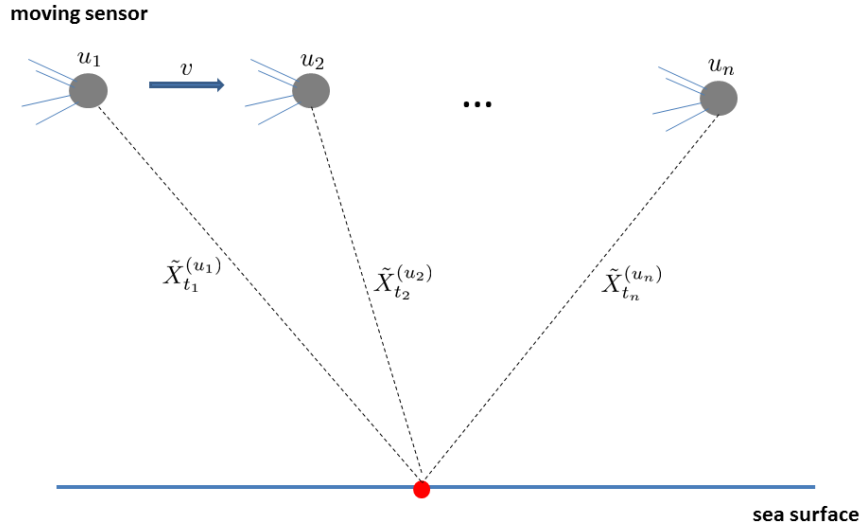
FIGURE 2 – Capteur mobile mesurant le clutter de mer $X_t^{(u)}$ aux positions $u_1, u_2, \ldots, u_n$ et temps $t_1, t_2, \ldots, t_n$. $\tilde{X}_t^{(u)} = X_t^{(u)}(\omega)$ est une réalisation du processus aléatoire.

Puisque le flou présent dans les images RSO de la mer s'explique par le mouvement de celle-ci, et donc par la variation de sa refléctivité pendant le processus d'intégration, nous proposons de synchroniser les différentes mesures grâces aux probabilités de transitions des différents processus aléatoires impliqués dans le clutter de mer. Si on considère par exemple

les deux premières mesures, on peut projeter la mesure au temps t_1 sur le temps t_2 grâce à la probabilité de transition. On obtient :

$$\tilde{X}_{t_1}^{(u_1)}, \tilde{X}_{t_2}^{(u_2)} \to p\left(X_{t_2}^{(u_1)} = x \mid X_{t_1}^{(u_1)} = \tilde{X}_{t_1}^{(u_1)}\right), \tilde{X}_{t_2}^{(u_2)}. \tag{12}$$

Nous avons posé et résolu les équations de Fokker-Planck associées respectivement à x_t et $\gamma_t^{(R)}$ (l'équation de Fokker-Planck de $\gamma_t^{(I)}$ est identique à celle de $\gamma_t^{(R)}$). Pour x_t, nous avons prouvé que les probabilités de transitions vers le futur sont :

$$p(x_t = x | x_0 = y) = \sum_{n=0}^{+\infty} \frac{\alpha L_n^{\alpha-1}(\alpha y) n!}{\Gamma(n+\alpha)} e^{-\mathcal{A}nt} e^{-\alpha x} (\alpha x)^{\alpha-1} L_n^{\alpha-1}(\alpha x), \tag{13}$$

où les $L_n^{\alpha-1}$ sont les polynômes de Laguerre et Γ la fonction gamma. Pour $\gamma_t^{(R)}$, nous avons prouvé que les probabilités de transitions vers le futur sont gaussiennes :

$$p\left(\gamma_t^{(R)} = x | \gamma_0^{(R)} = y\right) = \frac{1}{\sqrt{2\pi v(t)}} e^{-\frac{1}{2}\frac{(x-m_y(t))^2}{v(t)}}, \tag{14}$$

de moyenne :

$$m_y(t) = y e^{-\mathcal{B}t/2}, \tag{15}$$

et de variance :

$$v(t) = \frac{1 - e^{-\mathcal{B}t}}{2}. \tag{16}$$

Par homogénéité des processus x_t et γ_t, nous avons que $\forall h > 0$, $p(x_{t+h} = x | x_h = y) = p(x_t = x | x_0 = y)$ et que $p\left(\gamma_{t+h}^{(R)} = x | \gamma_h^{(R)} = y\right) = p\left(\gamma_t^{(R)} = x | \gamma_0^{(R)} = y\right)$. De plus, nous avons également montré que par application de la formule de Bayes, les probabilités de transitions vers le futur sont identiques à celles vers le passé, *i.e.* :

$$\begin{cases} p(x_{t-h} = y | x_t = x) = p(x_h = y | x_0 = x) \\ p(\gamma_{t-h}^{(R)} = y | \gamma_t^{(R)} = x) = p(\gamma_h^{(R)} = y | \gamma_0^{(R)} = x). \end{cases} \tag{17}$$

Nous avons réalisé des simulations numériques avec le schéma d'Euler-Maruyama. En simulant de nombreuses trajectoires partant de la même condition initiale déterministe, nous obtenons des probabilités de transition numériques que nous comparons aux probabilités de transitions analytiques. Des simulations sont représentées figure 3 pour $\gamma_t^{(R)}$, avec la condition initiale déterministe $\gamma_0^{(R)} = 2$. Les distributions analytiques et numériques concordent presque parfaitement. Nous observons la diffusion progressive des probabilités de transition (variance qui augmente), jusqu'à la distribution asymptotique. Ceci permet d'affirmer que la prédiction sur la valeur future du clutter de mer étant donnée sa valeur présente devient de plus en plus imprécise quand l'échéance de la prédiction s'éloigne. Des simulations et observations similaires ont été faites pour x_t.

5 Estimation des paramètres du clutter de mer

Nous avons noté dans la section 3 que le modèle de Field était paramétrisé par trois paramètres : $\mathcal{A}$ et α pour la SER x_t, et $\mathcal{B}$ pour le speckle γ_t. Nous nous intéressons ici à l'estimation de ces trois paramètres à partir de séries temporelles de x_t, $z_t = |\Psi_t|^2$ et γ_t. Comme montré ci-dessous, l'estimation de α se fait sous l'hypothèse que le clutter de mer est ergodique, tandis que celle de $\mathcal{A}$ et $\mathcal{B}$ se fait par maximum de vraisemblance.
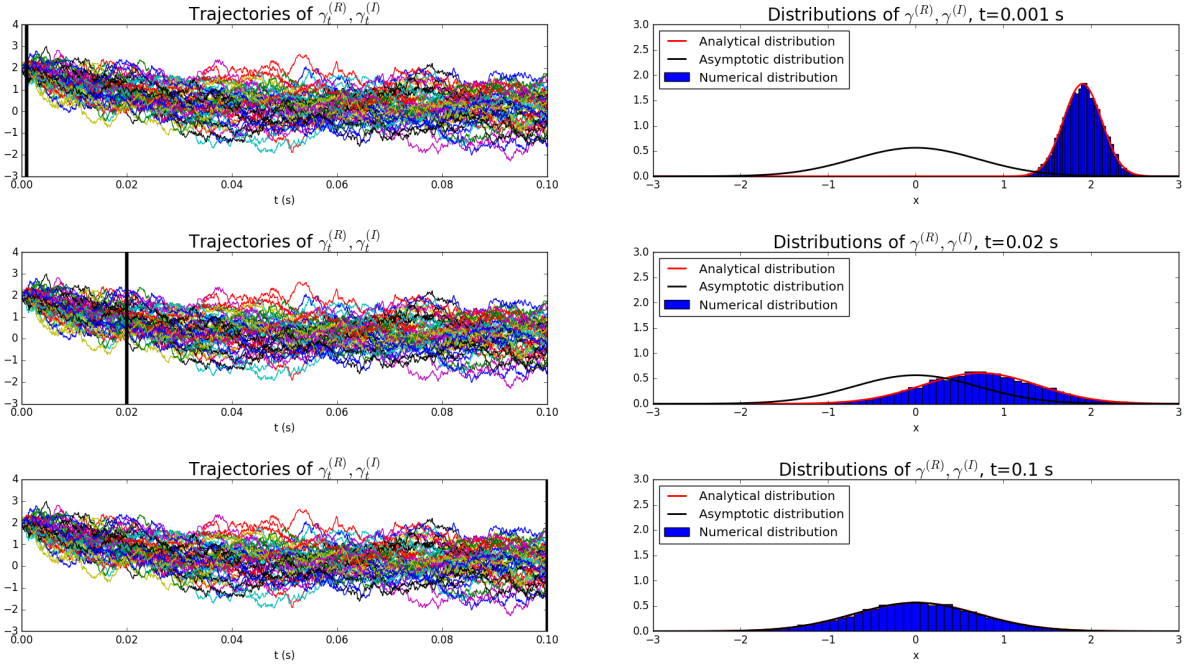
FIGURE 3 – Comparison entre les probabilités de transition analytique de $\gamma_t^{(R)}$ (équation (14)), et les probabilités de transitions numériques obtenues grâce au schéma d'Euler-Maruyama. 10000 trajectoires ont été simulées avec $\mathcal{B} = 100$ Hz et $y = 2$.

5.1 Estimation de α

Il est possible de montrer que la distribution stationnaire (voir section 2) de x_t est :

$$p(x_t = x) = \frac{\alpha^\alpha x^{\alpha-1} e^{-\alpha x}}{\Gamma(\alpha)}, \tag{18}$$

et que $\mathrm{var}(x_t) = 1/\alpha$ pour tout t. On suppose observer une série temporelle x_i aux temps $t_i = i\Delta t$, avec i allant de 0 à n. En supposant que le processus aléatoire x_t est ergodique, on obtient l'estimateur suivant pour α :

$$\tilde{\alpha}_x = \left[\frac{1}{n\Delta t} \sum_{i=0}^{n-1} (x_i - 1)^2 \Delta t \right]^{-1}. \tag{19}$$

On propose un autre estimateur de α à partir d'une série temporelle de l'intensité z_t. Pour ça, on note que dans le modèle de Field, la distribution stationnaire de z_t est :

$$p(z_t = x) = \frac{2b^{(\nu+1)/2} x^{(\nu-1)/2}}{\Gamma(\nu)} K_{\nu-1}(2\sqrt{bx}). \tag{20}$$

En utilisant les relations entre les moments d'une variable aléatoire K-distribué (voir [17] p 110), on obtient l'estimateur :

$$\tilde{\alpha}_z = \frac{2 \left(\sum_{i=0}^{n-1} z_i \Delta t \right)^2}{\sum_{i=0}^{n-1} (z_i - 1)^2 \Delta t - \left(\sum_{i=0}^{n-1} z_i \Delta t \right)^2}. \tag{21}$$

Nous comprendrons à la section 7 en quoi avoir les deux estimateurs $\tilde{\alpha}_x$ et $\tilde{\alpha}_z$ est utile.

5.2 Estimation de $\mathcal{A}$ et $\mathcal{B}$

Pour estimer $\mathcal{A}$ et $\mathcal{B}$, nous utilisons le principe du maximum de vraisemblance, qui consiste à maximiser la vraisemblance d'une série temporelle par rapport au paramètre recherché. Un processus aléatoire qui est solution d'une EDS (1) est un processus de Markov. Cette propriété peut être utilisée pour montrer que la vraisemblance (c'est à dire la distribution jointe) d'une série temporelle $\tilde{X}_{t_i}$ pour i allant de 0 à n s'écrit :

$$\mathcal{L} = p(X_{t_0} = \tilde{X}_0) \prod_{i=1}^{N} p(X_{t_i} = \tilde{X}_i \mid X_{t_{i-1}} = \tilde{X}_{i-1}). \tag{22}$$

Le premier terme $p(X_{t_0} = \tilde{X}_0)$ est la distribution stationnaire pour la première valeur observée, tandis que les autres correspondent aux différentes transitions. La vraisemblance est donc connue pour peu que la distribution stationnaire et les probabilités de transition le soient. Or, les probabilités de transition de x_t et $\gamma_t^{(R)}$ ont été calculées à la section 4. De plus, la distribution stationnaire de x_t est donnée par l'équation (18), et celle de $\gamma_t^{(R)}$ (ou $\gamma_t^{(I)}$) est simplement :

$$p\left(\gamma_t^{(R)} = x\right) = p\left(\gamma_t^{(I)} = x\right) = \frac{1}{\sqrt{\pi}}e^{-x^2}. \tag{23}$$

Pour estimer $\mathcal{A}$, on suppose α connu et on calcule la vraisemblance de la série temporelle $\tilde{x}_i$, $i = 0, 1, \ldots, n$. C'est une fonction de $\mathcal{A}$ qu'on maximise numériquement par rapport à $\mathcal{A}$ lorsque les probabilités de transition exactes (équation (13)) sont utilisées. En utilisant l'approximation du schéma d'Euler-Maruyama (voir section 2), on peut maximiser la vraisemblance analytiquement. Le paramètre estimé est alors solution du polynôme :

$$-\sum_{i=1}^{n} \frac{\alpha(\tilde{x}_{i-1} - 1)^2}{4\tilde{x}_{i-1}}\mathcal{A}^2 - \frac{n}{2}\mathcal{A} + \sum_{i=1}^{n} \frac{\alpha(\tilde{x}_i - \tilde{x}_{i-1})^2}{4\tilde{x}_{i-1}\Delta t} = 0. \tag{24}$$

Nous avons testé une autre approximation plus élaborée que le schéma d'Euler-Maruyama, que nous appelons approximation de Nowman (voir [12]) mais qui s'est révélée sans grand intérêt pour x_t. En effet, ses performances sont similaires à celles de l'approximation d'Euler-Maruyama mais sa complexité est supérieure.

Pour estimer $\mathcal{B}$, nous pouvons également utiliser les probabilités de transition exactes de $\gamma_t^{(R)}$ (équation (14)). Dans ce cas ci, elles sont strictement égales à l'approximation de Nowman, et la vraisemblance peut être maximisée analytiquement par rapport à $\mathcal{B}$. Le $\mathcal{B}$ estimé est solution du polynôme suivant :

$$\frac{n}{2}Y^3 - \sum_{i=1}^{n} \tilde{\gamma}_{i-1}^{(R)}\tilde{\gamma}_i^{(R)}Y^2 + \left(-\frac{n}{2} + \sum_{i=1}^{n}\left(\tilde{\gamma}_i^{(R)\,2} + \tilde{\gamma}_{i-1}^{(R)\,2}\right)\right)Y - \sum_{i=1}^{n} \tilde{\gamma}_{i-1}^{(R)}\tilde{\gamma}_i^{(R)} = 0, \tag{25}$$

tandis qu'en utilisant le schéma d'Euler-Maruyama, il est solution de :

$$-\sum_{i=1}^{n} \frac{\tilde{\gamma}_{i-1}^{(R)\,2}\Delta t}{4}\mathcal{B}^2 - \frac{n}{2}\mathcal{B} + \sum_{i=1}^{n} \frac{\left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)}\right)^2}{\Delta t} = 0. \tag{26}$$

Nous avons testé numériquement la performance de nos estimateurs. Par exemple, pour $\mathcal{A}$, nous avons considéré que sa valeur pouvait se situer entre 0.1 et 10 Hz. Pour chaque valeur dans $\{0.1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, nous avons généré $N = 1000$ trajectoires de la SER (avec

$\alpha = 1$), de durée 1 s chacune et avec $\Delta t = 0.001$ s. On obtient N estimées $\tilde{\mathcal{A}}_1, \tilde{\mathcal{A}}_2, \ldots, \tilde{\mathcal{A}}_N$, à partir desquelles le biais $\tilde{b}(\mathcal{A})$ et l'écart type d'estimation $\tilde{\sigma}(\mathcal{A})$ sont calculés. Les résultats sont représentés figure 4. On observe une croissance du biais et de l'écart type en fonction de $\mathcal{A}$, et que les probabilités de transitions exactes donnent des meilleurs résultats que les approximations. Un travail similaire a été fait pour $\mathcal{B}$, en considérant cette fois que sa valeur pouvait se situer entre 10 et 1000 Hz.

Malgré les meilleurs résultats des probabilités exactes observés à la figure 4, une analyse plus approfondie nous amène en fait à considérer que l'approximation d'Euler-Maruyama peut être utilisée de manière tout à fait satisfaisante, que ce soit pour $\mathcal{A}$ ou pour $\mathcal{B}$. En effet, la différence entre les estimées par les probabilités exactes et l'approximation d'Euler-Maruyama sont très faibles devant l'écart à la vraie valeur du paramètre. Il y a une très forte corrélation (trajectoire à trajectoire) de ces estimées, qui suggère qu'elles exploitent aussi bien l'une que l'autre l'information sur le paramètre contenue dans la trajectoire.

Dans le cadre d'un travail collaboratif avec Randolf Altmeyer (Université Humboldt de Berlin), nous avons proposé une méthode d'estimation alternative pour les paramètres $\mathcal{A}$ et $\mathcal{B}$, basées sur l'analyse de la volatilité intégrée [1] des processus x_t et $\gamma_t^{(R)}$. La méthode est significativement plus simple que le maximum de vraisemblance mais donne des résultats similaires. Ces travaux sont en annexe de la thèse.
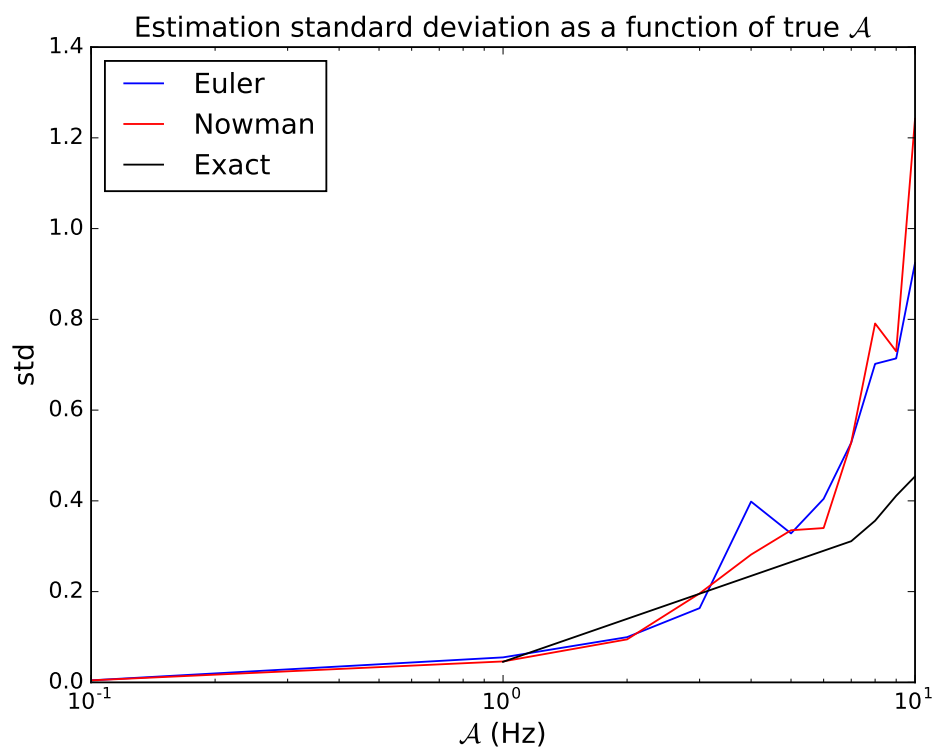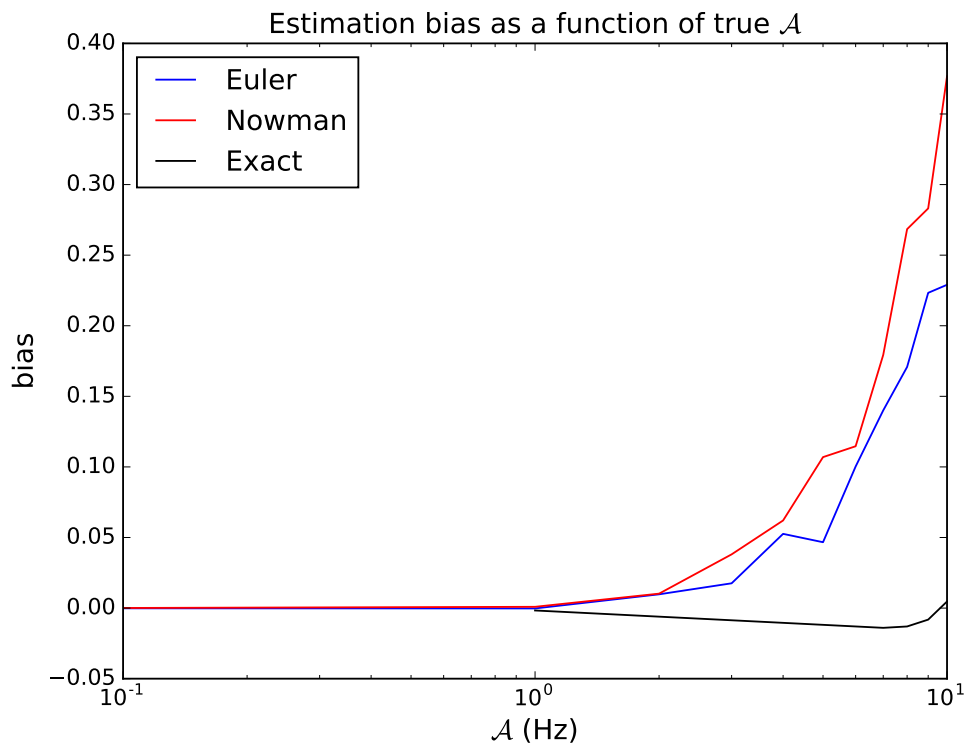
FIGURE 4 – Biais (haut) et écart type (bas) des estimateurs de $\mathcal{A}$ en fonction du vrai $\mathcal{A}$, calculés à partir de 1000 trajectoires de durée 1 s chacune et échantillonnées à 1000 Hz.

6 Prise en compte de la présence d'une cible

6.1 Modèle avec cible

Il est possible d'inclure la présence d'une cible dans le modèle de Field. Deux modèles ont été proposés dans [4]. Dans le modèle Homodyned K (HK), la réflectivité de la cible est une constante complexe ajoutée au clutter de mer. Si on dénote $\Psi_t^{(HK)}$ la réflectivité totale, on a :

$$\Psi_t^{(HK)} = \Psi_c^{(R)} + i\Psi_c^{(I)} + x_t^{1/2}\gamma_t, \tag{27}$$

où $\Psi_c = \Psi_c^{(R)} + i\Psi_c^{(I)}$ est la réflectivité de la cible. Le modèle Generalized K (GK), quant à lui, fait varier la réflectivité de la cible avec la SER x_t (see [4] p 71) :

$$\Psi_t^{(GK)} = \left(\Psi_c^{(R)} + i\Psi_c^{(I)}\right)\eta x_t + x_t^{1/2}\gamma_t, \tag{28}$$

où η est un facteur de couplage.

Nous avons montré dans les deux cas (HK et GK) qu'il était possible de réécrire les équations (10) du modèle de Field en prenant en compte la cible, et qu'on pouvait estimer la constante de cible Ψ_c par maximum de vraisemblance.

Dans le cas HK, en notant $\Psi_t^{(HK)} = R_t + iI_t$, nous avons montré en utilisant le calcul d'Itô que le processus multidimensionnel $\begin{bmatrix} x_t & R_t & I_t \end{bmatrix}^\top$ était solution de l'EDS :

$$
\begin{bmatrix} \mathrm{d}x_t \\ \mathrm{d}R_t \\ \mathrm{d}I_t \end{bmatrix} = \begin{bmatrix} \mathcal{A}(1-x_t) \\ -\frac{\mathcal{A}+\mathcal{B}}{2}\left(R_t - \Psi_c^{(R)}\right) + \frac{\mathcal{A}\left(R_t - \Psi_c^{(R)}\right)}{2x_t}\left(1-\frac{1}{2\alpha}\right) \\ -\frac{\mathcal{A}+\mathcal{B}}{2}\left(I_t - \Psi_c^{(I)}\right) + \frac{\mathcal{A}\left(I_t - \Psi_c^{(I)}\right)}{2x_t}\left(1-\frac{1}{2\alpha}\right) \end{bmatrix}\mathrm{d}t
$$
$$
+ \begin{bmatrix} \left(\frac{2\mathcal{A}x_t}{\alpha}\right)^{1/2} & 0 & 0 \\ \frac{\left(R_t - \Psi_c^{(R)}\right)}{x_t^{1/2}}\left(\frac{\mathcal{A}}{2\alpha}\right)^{1/2} & \left(\frac{\mathcal{B}x_t}{2}\right)^{1/2} & 0 \\ \frac{\left(I_t - \Psi_c^{(I)}\right)}{x_t^{1/2}}\left(\frac{\mathcal{A}}{2\alpha}\right)^{1/2} & 0 & \left(\frac{\mathcal{B}x_t}{2}\right)^{1/2} \end{bmatrix}\begin{bmatrix} \mathrm{d}W_t^{(x)} \\ \mathrm{d}W_t^{(R)} \\ \mathrm{d}W_t^{(I)} \end{bmatrix}. \tag{29}
$$

Dans le cas GK, en notant de nouveau $\Psi_t^{(GK)} = R_t + iI_t$, nous avons montré que :

$$
\begin{bmatrix} \mathrm{d}x_t \\ \mathrm{d}R_t \\ \mathrm{d}I_t \end{bmatrix} = \begin{bmatrix} \mathcal{A}(1-x_t) \\ \eta\Psi_c^{(R)}\mathcal{A}(1-x_t) + \left(R_t - \Psi_c^{(R)}\eta x_t\right)\left(-\frac{\mathcal{A}+\mathcal{B}}{2} + \frac{\mathcal{A}}{2x_t}\left(1-\frac{1}{2\alpha}\right)\right) \\ \eta\Psi_c^{(I)}\mathcal{A}(1-x_t) + \left(I_t - \Psi_c^{(I)}\eta x_t\right)\left(-\frac{\mathcal{A}+\mathcal{B}}{2} + \frac{\mathcal{A}}{2x_t}\left(1-\frac{1}{2\alpha}\right)\right) \end{bmatrix}\mathrm{d}t
$$
$$
+ \begin{bmatrix} \left(\frac{2\mathcal{A}x_t}{\alpha}\right)^{1/2} & 0 & 0 \\ \frac{R_t - \Psi_c^{(R)}\eta x_t}{x_t^{1/2}}\left(\frac{\mathcal{A}}{2\alpha}\right)^{1/2} + \eta\Psi_c^{(R)}\left(\frac{2\mathcal{A}x_t}{\alpha}\right)^{1/2} & \left(\frac{\mathcal{B}x_t}{2}\right)^{1/2} & 0 \\ \frac{I_t - \Psi_c^{(I)}\eta x_t}{x_t^{1/2}}\left(\frac{\mathcal{A}}{2\alpha}\right)^{1/2} + \eta\Psi_c^{(I)}\left(\frac{2\mathcal{A}x_t}{\alpha}\right)^{1/2} & 0 & \left(\frac{\mathcal{B}x_t}{2}\right)^{1/2} \end{bmatrix}\begin{bmatrix} \mathrm{d}W_t^{(x)} \\ \mathrm{d}W_t^{(R)} \\ \mathrm{d}W_t^{(I)} \end{bmatrix}. \tag{30}
$$

6.2 Estimation des paramètres de cible

Pour estimer le paramètre complexe Ψ_c, nous avons supposé que trois séries temporelles étaient observées :

$$(\tilde{x}, \tilde{R}, \tilde{I}) = \left\{\left(\tilde{x}_k, \tilde{R}_k, \tilde{I}_k\right), k = 0, 1, \ldots, n\right\},$$

où les mesures sont faites aux temps t_k et où $\forall k, \ t_k - t_{k-1} = \Delta t$ est une constante. La vraisemblance de $(\tilde{x}, \tilde{R}, \tilde{I})$ est alors :

$$\mathcal{L}\left(\tilde{x}, \tilde{R}, \tilde{I}; \Psi_c\right) = p_{\Psi_c}^\infty \prod_{k=1}^n p_{\Psi_c}^{(k)}, \tag{31}$$

avec

$$\begin{aligned} p_{\Psi_c}^\infty &= p_{\Psi_c}((x_{t_0}, R_{t_0}, I_{t_0}) = (\tilde{x}_0, \tilde{R}_0, \tilde{I}_0)) \\ &= p_{\Psi_c}^\infty(\tilde{x}_0, \tilde{R}_0, \tilde{I}_0) \end{aligned} \tag{32}$$

et

$$p_{\Psi_c}^{(k)} = p_{\Psi_c}\left((x_{t_k}, R_{t_k}, I_{t_k}) = (\tilde{x}_k, \tilde{R}_k, \tilde{I}_k) \mid (x_{t_{k-1}}, R_{t_{k-1}}, I_{t_{k-1}}) = (\tilde{x}_{k-1}, \tilde{R}_{k-1}, \tilde{I}_{k-1})\right). \tag{33}$$

Les distributions stationnaires/asymptotiques $p_{\Psi_c}^\infty$ peuvent être obtenues assez simplement à partir de celles de $\gamma_t^{(R)}$, $\gamma_t^{(I)}$, x_t, et en exploitant l'indépendance de $\gamma_t^{(R)}$ et $\gamma_t^{(I)}$. Pour obtenir les probabilités de transition $p_{\Psi_c}^{(k)}$, nous utilisons de nouveau le schéma d'Euler-Maruyama, qui permet d'approximer les probabilités de transition du processus $(\tilde{x}, \tilde{R}, \tilde{I})$ par des gaussiennes multidimensionnelles.

Nous avons montré qu'il était alors possible de maximiser la vraisemblance relativement à Ψ_c. Dans le cas HK, on obtient des estimateurs de la forme :

$$\begin{cases} \tilde{\Psi}_{c,ML}^{(R)} = f_{HK}(\tilde{x}, \tilde{R}) \\ \tilde{\Psi}_{c,ML}^{(I)} = f_{HK}(\tilde{x}, \tilde{I}), \end{cases} \tag{34}$$

et dans le cas GK :

$$\begin{cases} \tilde{\Psi}_{c,ML}^{(R)} = f_{GK}(\tilde{x}, \tilde{R}) \\ \tilde{\Psi}_{c,ML}^{(I)} = f_{GK}(\tilde{x}, \tilde{I}). \end{cases} \tag{35}$$

Nous avons comparé numériquement les résultats des estimateurs par maximum de vraisemblance avec ceux d'estimateurs beaucoup plus simple basé sur la propriété $\mathbb{E}[\Psi_t] = 0$ et supposant l'ergodicité de Ψ_t :

$$\begin{cases} \tilde{\Psi}_{c,e} = \frac{1}{n\Delta t} \sum_{i=0}^{n-1} \tilde{\Psi}_i^{(HK)} \Delta t & (HK) \\ \tilde{\Psi}_{c,e} = \frac{1}{n\Delta t} \sum_{i=0}^{n-1} \tilde{\Psi}_i^{(GK)} \Delta t & (GK). \end{cases} \tag{36}$$

Certains résultats sont représentés figure 5. Dans les deux cas, HK et GK, nous avons simulé 1000 trajectoires de 1 s avec $\Delta t = 0.001$ s en incluant une cible $\Psi_c = \Psi_c^{(R)} = \sqrt{10}$, et estimé $\Psi_c^{(R)}$ en utilisant les deux estimateurs (maximum de vraisemblance et ergodicité). La figure 5, représente sous forme d'un "scatter plot" les différentes estimations en fonction du numéro de trajectoires. Nous voyons visuellement que l'estimateur par maximum de vraisemblance n'est pas meilleur que celui basé sur l'ergodicité pour le cas HK, tandis qu'il est bien meilleur dans le cas GK, à ceci près qu'il génère quelques données aberrantes.

Dans le cadre de notre travail collaboratif avec Randolf Altmeyer (Université Humboldt de Berlin), nous avons proposé une méthode d'estimation de Ψ_c alternative, basées sur l'analyse de la volatilité intégrée [1] de $\begin{bmatrix} x_t & R_t & I_t \end{bmatrix}^\top$. La méthode est significativement plus simple que le maximum de vraisemblance. Ses résultats sont équivalents à ceux du maximum de vraisemblance dans le cas GK, mais moindres dans le cas HK. Ces travaux sont en annexe de la thèse.
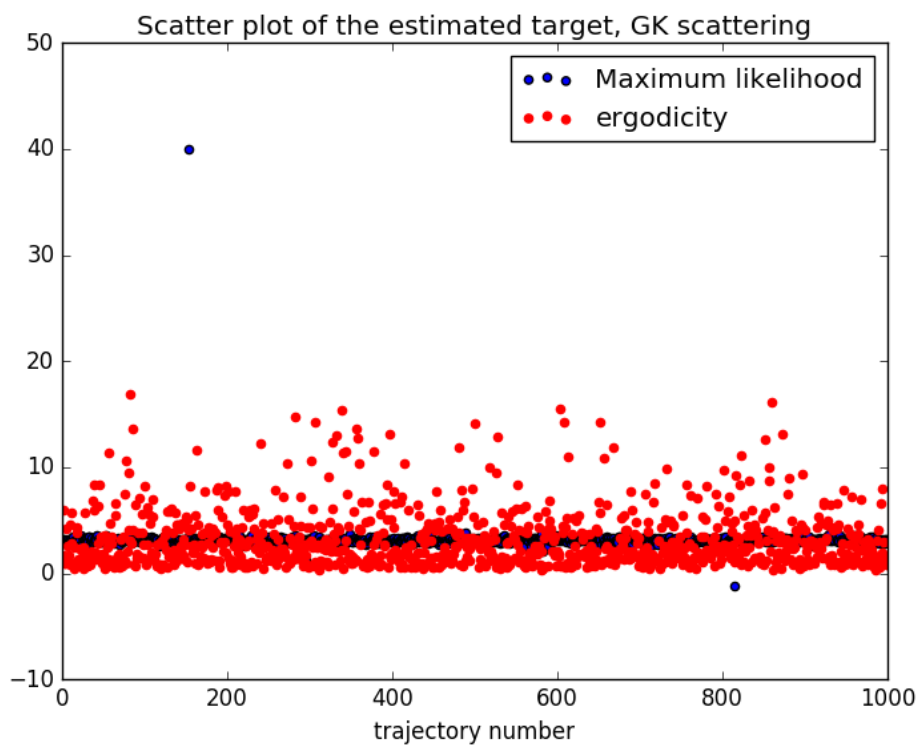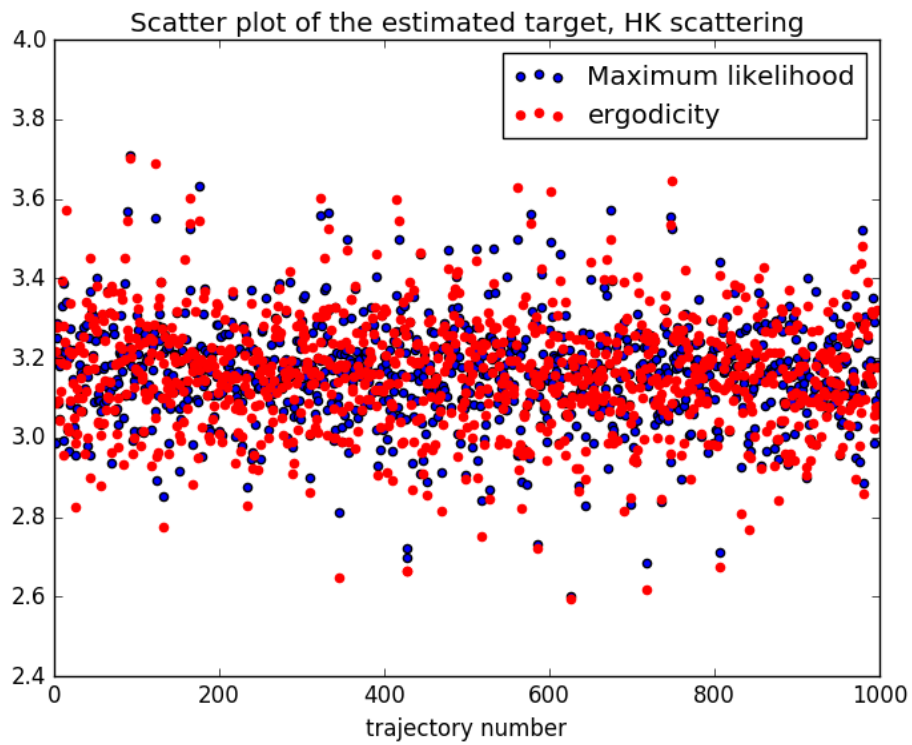
Figure 5 – Comparaison entre les estimateurs par maximum de vraisemblance et basés sur l'hypothèse d'ergodicité pour le cas HK (haut) et GK (bas).

7 Non-observabilité de x_t

Nous avons présenté le modèle de Field dans la section 3 et expliqué que la refléctivité complexe de la surface de la mer est décrite grâce aux trois processus x_t, $\gamma_t^{(R)}$ et $\gamma_t^{(I)}$ *via* l'équation :

$$\Psi_t = x_t^{1/2}\left(\gamma_t^{(R)} + i\gamma_t^{(I)}\right). \tag{37}$$

Dans les sections 4 à 6, nous avons supposé que les séries temporelles de x_t, $\gamma_t^{(R)}$ et $\gamma_t^{(I)}$ étaient observées. En réalité, un radar cohérent ne peut pas mesurer ces processus, mais uniquement Ψ_t, et donc $R_t = x_t^{1/2}\gamma_t^{(R)}$ et $I_t = x_t^{1/2}\gamma_t^{(I)}$ (parties réelles et imaginaires). Afin de répondre à ce problème, nous avons proposé des estimateurs de la SER x_t basés sur Ψ_t en utilisant le fait que x_t évolue lentement par rapport à γ_t ($\mathcal{A} \ll \mathcal{B}$). Nous avons ensuite proposé une séquence d'estimateurs pour la SER, le speckle, et les paramètres du clutter de mer que nous avons analysée numériquement.

Pour estimer x_t, nous considérons d'abord que t est fixé et qu'une série temporelle $\{\Psi_{t_k}, k = 0, 1, \ldots, n\}$, telle que $t_k - t_{k-1} = \Delta t$ pour tout k, est observée. On considère une fenêtre Δ_t contenant t (de préférence centrée en t) et on suppose que pour tout $k \in \Delta_t$ (tout k tel que $t_k \in \Delta_t$), on a $x_{t_k} \approx x_t$. En posant $\Delta_k\Psi = \Psi_{t_k} - \Psi_{t_{k-1}}$, nous avons proposé l'estimateur suivant pour x_t :

$$\bar{x}_t = \frac{1}{\mathcal{B}\Delta t N}\sum_{k \in \Delta_t}|\Delta_k\Psi|^2, \tag{38}$$

où N est le nombre de timesteps dans la fenêtre Δ_t, c'est à dire le nombre d'incréments $|\Delta_k\Psi|$ moyennés pour calculer $\bar{x}_t$. En faisant glisser la fenêtre Δ_t, on peut estimer x_t pour tout t. Le problème de l'estimateur $\bar{x}_t$ est qu'il n'est pas calculable en pratique parce que $\mathcal{B}$ n'est pas connu. Pour estimer $\mathcal{B}$, on doit proposer un autre estimateur que le maximum de vraisemblance $\tilde{B}_{ML}$ obtenu section 5 puisque x_t, donc γ_t, n'est pas observé.

Nous avons montré qu'il était pertinent de proposer l'estimateur suivant pour $\mathcal{B}$:

$$\tilde{\mathcal{B}}_\Psi = \frac{1}{m\Delta t N}\sum_{i=1}^{m}\sum_{k \in \Delta_i}|\Delta_k\Psi|^2, \tag{39}$$

où quelque soit i, Δ_i est une fenêtre centrée en t_i et qui contient N incréments, et où m est le nombre de fenêtres. $\tilde{\mathcal{B}}_\Psi$ est calculable directement à partir de données observables, contrairement à $\tilde{B}_{ML}$. On obtient l'estimateur suivant (calculable sans connaître $\mathcal{B}$) :

$$\breve{x}_t = \frac{1}{\tilde{\mathcal{B}}_\Psi\Delta t N}\sum_{k \in \Delta}|\Delta_k\Psi|^2. \tag{40}$$

Il s'ensuit l'estimateur suivant pour γ_t :

$$\breve{\gamma}_t = \frac{\Psi_t}{\breve{x}_t^{1/2}}. \tag{41}$$

Le seul degré de liberté laissé à l'utilisateur pour calculer $\breve{x}_t$ est N, le nombre d'incréments par fenêtres. En s'appuyant sur [3] et sur des simulations numériques, nous avons montré que

le nombre optimal d'incréments (au sens de la minimisation de l'erreur moyenne quadratique) pouvait être approché par la formule suivante :

$$N_{opt} = 0.64 \left(\frac{12(\alpha + 1)}{\mathcal{A}\Delta t} \right)^{1/2}, \tag{42}$$

qui dépend de $\mathcal{A}$ et α (inconnus, à estimer) et de Δt (connu).

Nous avons également proposé un autre estimateur de x_t, plus simple et basé sur la série temporelle d'intensité $\{z_{t_k}, k = 0, 1, \ldots, n\}$, avec $z_{t_k} = |\Psi_{t_k}|^2$ pour tout k :

$$\hat{x}_t = \frac{1}{N} \sum_{k \in \Delta_t} z_{t_k}. \tag{43}$$

La figure 6 représente un exemple de trajectoire de x_t simulée numériquement, comparée aux trois estimateurs proposés. Nous observons que $\bar{x}_t$ et $\breve{x}_t$ sont meilleurs que $\hat{x}_t$. Puisque seul $\breve{x}_t$ est calculable en pratique (ici nous connaissions le vrai $\mathcal{B}$ puisque la trajectoire a été simulée), nous ne retenons que cet estimateur.
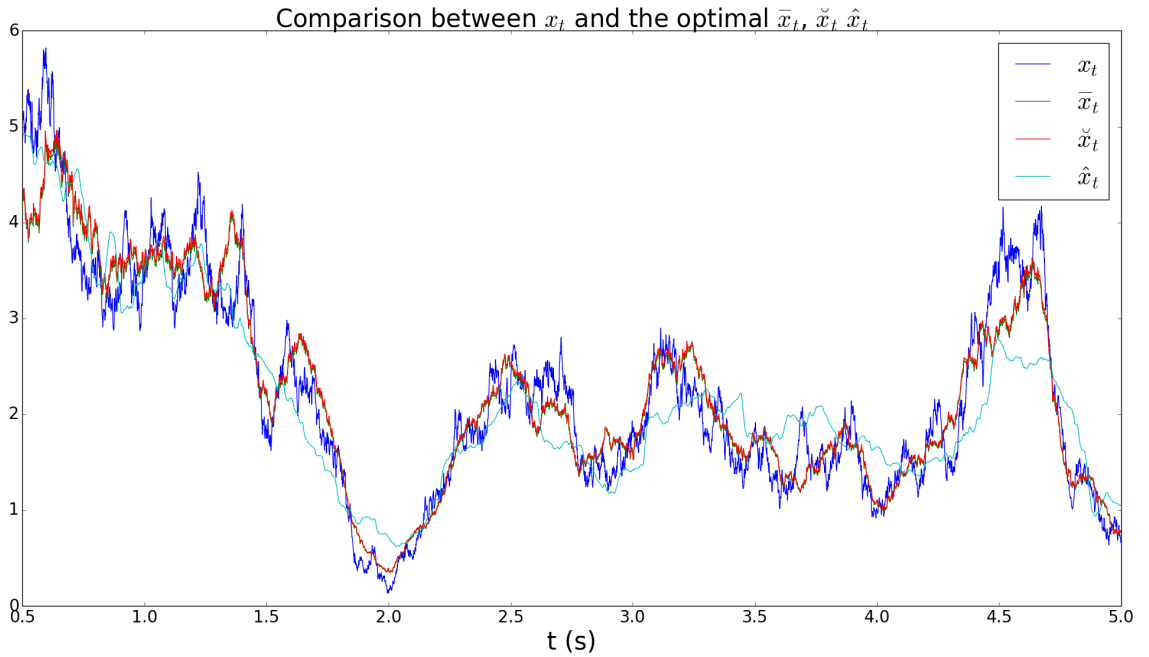
FIGURE 6 – Exemple de trajectoire de x_t et de ses trois estimations $\bar{x}_t$, $\breve{x}_t$ et $\hat{x}_t$.

Bien que $\breve{x}_t$ et $\tilde{B}_\Psi$ soient calculables à partir de Ψ_t, il existe toujours un couplage (circularité) entre l'estimation de x_t et celles de $\mathcal{A}$ et α, en raison du fait que la taille des fenêtres N_{opt} dépend de $\mathcal{A}$ et α (à estimer). Nous proposons d'utiliser dans un premier temps l'estimateur $\tilde{\alpha}_z$ (equation (21)) pour α (calculable à partir de données observées) et la valeur par défaut $\mathcal{A} = 5$ Hz pour calculer N_{opt}. Des premières estimations de x_t et γ_t sont alors calculées avec les équations (40) et (41). A partir de $\breve{x}_t$, il est possible d'obtenir des estimations plus fines de $\mathcal{A}$ et α en calculant respectivement $\tilde{\mathcal{A}}_{ML}$ (estimateur par maximum de vraisemblance, section 5) et $\tilde{\alpha}_x$. On peut alors recalculer N_{opt} avec $\tilde{\mathcal{A}}_{ML}$ et $\tilde{\alpha}_x$, recalculer $\breve{x}_t$ et $\breve{\gamma}_t$ et réitérer. La séquence d'estimateurs correspondante est représentée figure 7.

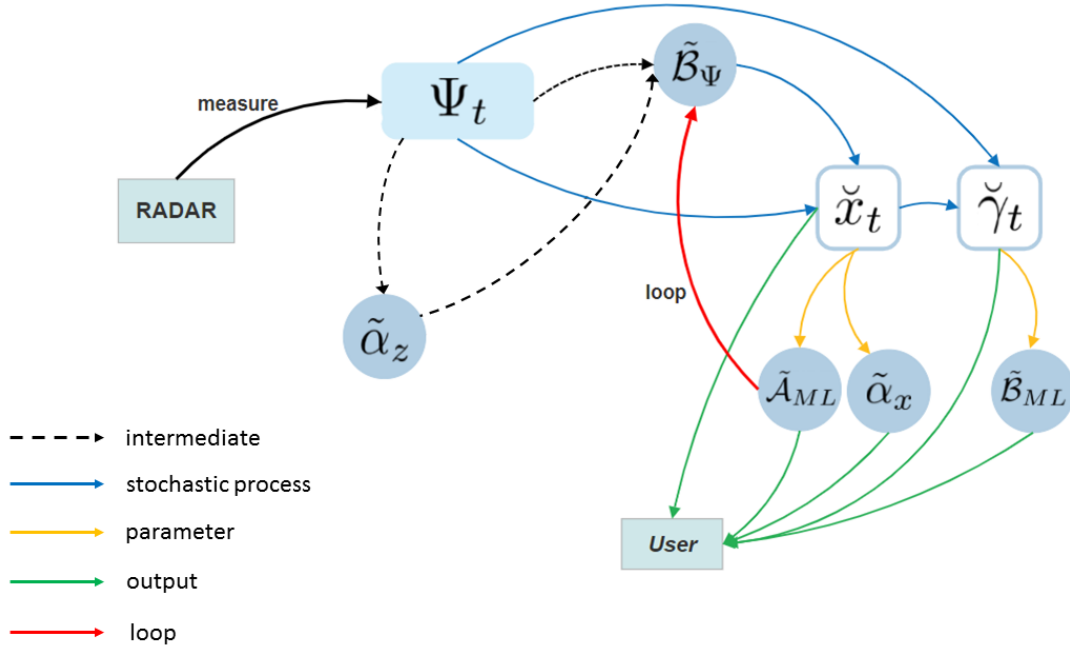FIGURE 7 – Séquence d'estimateurs pour estimer $\mathcal{A}$, $\mathcal{B}$, α, x_t et γ_t.

Nous avons testé numériquement cette séquence d'estimateurs sur des trajectoires de Ψ_t simulées avec des paramètres connus, et observé certains résultats qui ont nécessité une analyse approfondie. D'une part, nous avons remarqué qu'il était nécessaire de sous-échantillonner $\breve{x}_t$ avant de calculer $\tilde{\mathcal{A}}_{ML}$, en raison d'un phénomène de lissage lié à l'utilisation d'une fenêtre glissante pour le calcul de $\breve{x}_t$. D'autre part, la qualité de l'estimation ne peut pas uniquement être mesurée à la capacité des estimateurs à retrouver les vrais paramètres utilisés pour faire les simulations. Les estimateurs par maximum de vraisemblance sont par définition optimaux dans le sens où ils maximisent la fonction de vraisemblance de $\breve{x}_t$ et $\breve{\gamma}_t$, et expliquent donc mieux les trajectoires observées en termes de probabilités de transition que n'importe quels autres estimateurs (y compris les vrais paramètres). Cette conclusion est consubstantielle à la non-observabilité de la factorisation de Ψ_t par x_t et γ_t, et à la subjectivité de sa factorisation par $\breve{x}_t$ et $\breve{\gamma}_t$.

8 Conclusion

La télédétection en milieu maritime doit inévitablement s'accomoder de la présence du signal réflechi par la mer, le clutter de mer, qui s'apparente à du bruit. Une connaissance fine de celui-ci est souhaitable quelque soit la finalité (détection, imagerie etc). Dans cette thèse, nous avons étudié le modèle de Field [4], qui a l'avantage de décrire le clutter de mer en termes de densités de probabilités, comme la K distribution qui était préalablement connue, tout en modélisant la dynamique du processus grâce au formalisme des équations différentielles stochastiques.

Dans un premier temps, nous avons introduit les notions d'équation différentielle stochastique, d'équation de Fokker-Planck et de probabilités de transition. Nous avons ensuite

présenté le modèle de Field en expliquant qu'il dérive du modèle de la marche aléatoire (population de diffuseurs discrets) mais prend en compte la dimension temporelle grâce à un modèle de population pour le nombre de diffuseurs et à des équations différentielles stochastiques pour la phase des diffuseurs.

Nous avons posé et résolu les équations de Fokker-Planck de x_t et de γ_t pour calculer leurs probabilités de transition vers le futur et le passé. Nous avons suggéré qu'il était possible d'utiliser ces probabilités de transition pour synchroniser des observations faites à des positions et temps différents, comme c'est le cas dans l'imagerie Radar à Synthèse d'Ouverture (RSO). Ces résultats ont été publiés [15].

Nous avons montré qu'il était possible d'estimer les paramètres du modèle de Field, $\mathcal{A}$, $\mathcal{B}$ et α, à partir de séries temporelles de x_t pour $\mathcal{A}$ et α, et de γ_t pour $\mathcal{B}$. Nous avons obtenu deux estimateurs pour α, qui exploitent l'hypothèse d'ergodicité du clutter de mer, et des estimateurs par maximum de vraisemblance pour $\mathcal{A}$ et $\mathcal{B}$. Ces résultats ont été également publiés [14].

Nous avons ensuite pris en compte la présence d'une cible, et montré qu'il était possible d'estimer les paramètres de la cible par maximum de vraisemblance en utilisant le schéma d'Euler-Maruyama pour approximer les probabilités de transitions présentes dans la fonction de vraisemblance. Nous avons synthétisé ces travaux sous la forme d'un article soumis à *Waves in Random and Complex Media* [16].

Enfin, et dans un souci d'applicabilité des résultats précédents, nous avons travaillé sur la non-observabilité de x_t. Nous avons construit un estimateur $\breve{x}_t$ de x_t, et finalement proposé une chaine d'estimateurs pour x_t, γ_t, $\mathcal{A}$, $\mathcal{B}$ et α.

Le modèle de Field est éminemment théorique et difficilement exploitable dans sa présentation initiale [4], en raison de son caractère particulièrement synthétique. Nous pensons que nos travaux ont contribué d'une part à éclaircir le modèle en le présentant différemment, d'autre part à le rendre applicable dans le cadre du traitement de séries temporelles du clutter de mer. Nous avons vu que l'estimation des paramètres du modèle n'était possible qu'après avoir estimé la SER x_t, qui est non-observable. Un des résultats les plus notables de notre travail est la séquence d'estimateurs de la figure 7, qui, malgré sa perfectibilité, est applicable telle quelle à des données réelles. Nous avons remarqué qu'il y avait une incontournable subjectivité dans la factorisation du clutter de mer par la SER et le speckle, mais que celle-ci n'était pas forcément problématique. Ce qui est recherché est en effet surtout une description précise des données sur lesquelles nous sommes susceptibles de travailler en réalité, à savoir Ψ_t, $\breve{x}_t$ et $\breve{\gamma}_t$. Néanmoins, une mauvaise factorisation pourrait s'avérer délétère pour simuler numériquement des trajectoires ressemblant aux données observées. Le problème de la non-observabilité de x_t mériterait une étude plus poussée.

L'utilisation du modèle de Field pour l'imagerie RSO reste à explorer. Elle nécessiterait un focus particulier sur les algorithmes RSO, mais pourrait s'avérer difficile à mettre en pratique en particulier pour l'estimation des paramètres (absence de séries temporelles longues d'une même zone due au mouvement du capteur). Dans ce contexte, l'estimation des paramètres du clutter de mer devrait être repensée en n'exploitant pas seulement l'information temporelle, mais également spatiale. Nous pensons cependant qu'il est possible à présent d'exploiter le modèle de Field dans une optique de détection d'anomalies, par comparaison des statistiques du signal observé avec celles qui sont attendues en l'absence d'anomalie.

Enfin, il faut souligner les limites inhérentes au modèle de Field et qu'on est susceptible de rencontrer dans certains cas. D'une part, une des hypothèses clés est que le nombre de

diffuseurs contribuant à la réflectivité de la mer est infini. Cette limite peut être atteinte dans le cas de radars très haute résolution, où l'apparition d'un seul diffuseur supplémentaire peut être confondue avec celle d'une cible. D'autre part, l'hypothèse de mouvement brownien des diffuseurs et de leur indépendance peut aussi être questionnée. Elle suppose en effet que le mouvement de la surface de la mer n'est pas structuré au sein de la cellule de résolution, ce qui n'est pas acceptable par exemple en cas de forte houle et encore une fois de radars haute résolution.

Références

[1] Y. Aït-Sahalia and J. Jacod. *High-Frequency Financial Econometrics*. Princeton University Press, 2014.

[2] Richard Bamler and Birgit Schättler. SAR Data Acquisition and Image Formation, 1993.

[3] P. Fayard and T. R. Field. Optimal inference of the scattering cross-section through the phase decoherence. *Waves in Random and Complex Media*, 18(4) :571–584, 2008.

[4] T. R. Field. *Electromagnetic Scattering from Random Media*. Oxford University Press, 2009.

[5] E. Jakeman and K. D. Ridley. *Modeling Fluctuations in Scattered Waves*. CRC Press : Series in Optics and Optoelectronics, 2006.

[6] Samuel W. McCandless Jr. and Christopher R. Jackson. *Synthetic Aperture Radar Marine User's Manual, chapter 1*. http ://www.sarusersmanual.com/.

[7] M. B. Kanevsky. *Radar Imaging of the Ocean Waves*. Elsevier, 2008.

[8] J.-C. Laleuf. *Processus et intégrales stochastiques*. Editions Ellipses, 2014.

[9] H. Maitre. *Processing of Synthetic Aperture Radar (SAR) Images*. Wiley-IEEE Press, 1 edition, 2008.

[10] D. Massonet and J-C. Souyris. *Imaging with Synthetic Aperture Radar*. EPFL Press, 2008.

[11] B. Oksendal. *Stochastic Differential Equations : An Introduction with Applications, Fifth Edition*. Springer-Verlag, 2000.

[12] P. C. B. Phillips and J. Yu. Maximum Likelihood and Gaussian Estimation of Continuous Time Models in Finance. *Handbook of Financial Time Series*, 2009.

[13] H. Risken. *The Fokker-Planck Equation : methods of solution and applications*. Springer, 1989.

[14] C. J. Roussel, A. Coatanhay, and A. Baussard. Estimation of the parameters of stochastic differential equations for sea clutter. *IET Radar, Sonar and Navigation*, 13(4) :497 − 504, 2018.

[15] C. J. Roussel, A. Coatanhay, and A. Baussard. Forward and backward probabilistic inference of the sea clutter. *Waves in Random and Complex Media*, 29(3) :540–568, 2018.

[16] C. J. Roussel, A. Coatanhay, and A. Baussard. Detection of a coherent scatterer in a random medium : an approach based on transition probabilities. *submitted to Waves in Random and Complex Media*, 2019.

[17] K. D. Ward, R. J. A. Tough, and S. Watts. *Sea Clutter : Scattering, the K distribution and Radar Performance*. 20. IET Radar, Sonar and Navigation, first edition edition, 2006.

Contents

Introduction

It has always been both fascinating and scary, both a hindrance and a lever, an impassable wall or a semi-infinite space to escape. The sea covers two thirds of planet Earth. Having a long coastline and a large maritime territory is bittersweet: on one hand, it means that you are potentially very wealthy, but on the other hand, it is likely that some people will be willing to prey on your wealth or have bad intentions in general. Besides these sovereignty matters, we now also have to face environmentally unfriendly entities. Some countries, like France, are particularly concerned. Indeed, our maritime territory is almost twenty times larger than our land, which is enormous and comes with responsabilities: that of watching this territory.

The general topic of our research is remote sensing in a maritime environment. Maritime surveillance, which we have just put forward, is a good example of a motivated application related to that topic. We can cite other ones like air space surveillance from a boat, oceanography, or sea ice monitoring. Amongst the possible types of sensors, we are interested in active radars (Radio Detection and Ranging). Radars emit electromagnetic waves with wavelengths which can range from a few millimeters to tens of meters depending on the type of radar. The radar waves interact with the environment (where the meaning is no longer ecological but refer to the surroundings of the radar), which scatters part of the waves in all directions. If a receiver located somewhere records the scattered waves in form of a digital signal, it is possible to process the signal in order to learn about the environment by inversion. Some radars are also able to produce images. If the RMS Titanic had been equipped with a radar (which was not possible since they have been mostly invented during World War II), it may have detected the iceberg and dodged it, saving thousands of lives in 1912, and millions of tears in 1997. The advantage of radars over *e.g.* optical sensors is that radar waves can penetrate through clouds (relevant if emitted from an airplane or a satellite). That the system is active is also an advantage since it can work day and night.

The particularity of radar remote sensing in a maritime environment, is that the sea surface receives and scatters part of the emitted radar waves (radar waves cannot penetrate deep into water). The contribution of the sea is therefore inevitably present in the recorded signals. If it is the sea itself which is studied, this is of course desirable. However, if the aim is to detect a boat, the sea surface contribution should ideally be filtered out.

The sea surface, despite some structure, is complex, dynamic, chaotic, unpredictable. It has been shown that the radar echo from the sea surface, the reflectivity, has a fractal dimension proportional to that of the sea surface [23]. This is a demonstration of the fact that the dynamic nature and unpredictability of the sea surface is transmitted to its reflectivity. For this reason, the sea surface reflectivity is called sea clutter. The sea clutter carries information about the sea surface, which can be used to estimate the wind intensity, or sea surface waves with long wavelengths [57]. However, it is noise in the detection problem. Since *"we must free ourselves of the hope that the sea will ever rest"*, we have to cope with

the presence of sea clutter and develop signal processing algorithms accordingly. To do so, a good understanding and characterization of its features is necessary.

A first approach is based on electromagnetic theory (starting from Maxwell's equations) and aim at giving analytical expressions for the radar waves scattered by the sea surface, in terms of integrals over the surface. Traditionally, it leads to the Stratton-Chu equations [132], [142], which is a coupled system of integral equations for the electric and magnetic field. In practice, the system cannot be solved analytically for the sea surface. One either has to use numerical methods to solve the system [124], or to make approximations to simplify the equations, which leads to asymptotic models [113] like the Kirchoff (tangent plane) approximation. However, from a signal processing point of view, both methods are quite limited. On one hand, numerical methods require very large computational power and time to solve the equations and obtain time series of sea clutter. On the other hand, asymptotic models usually include a statistical model for the sea surface (for example a power spectrum), which represents the sea surface at fixed time but not its motion. Consequently, the temporal correlation of the sea clutter is lost.

A second approach is based on fractal analysis of sea clutter. It summarizes the complexity of sea clutter with mainly one quantity: the fractal dimension, which has the merit of being computable directly from time series. For example, it is shown in [23] that under some conditions, the fractal dimension of sea clutter is equal to the fractal dimension of the sea surface minus one. To our best knowledge, [90] was one of the first papers on the subject of fractal analysis of sea clutter, and it proposed to use it for target detection. For more on the subject, see *e.g.* [90], [23], [24], [96], [91], [74]. Fractal analysis certainly has valuable information to provide, but one of its drawback it that it does not have predictive power.

A third approach is statistical analysis and describes the sea clutter in terms of probability distributions. Statistical models have the advantage of offering (some) predictability power and lead quickly to target detection (see [141]). A first model is the Gaussian model for sea clutter (also called Rayleigh). It can be derived analytically by considering that the sea surface is approximated by a constant population of discrete scatterers. This is known as the random walk model [70], here with fixed "step number" (*i.e.* number of scatterers in the population). Under some assumptions, it can be proven that, as the number of scatterers goes to infinity, the joint distribution of the real and imaginary parts of the sea clutter becomes Gaussian, the modulus of the clutter becomes Rayleigh-distributed and the squared modulus of the clutter becomes exponential-distributed. It was soon observed that the Gaussian model was a bit simplistic, and more advanced models were proposed. One has particularly gained respectability: the K distribution [141], which is a probability distribution for the squared modulus of the sea surface complex reflectivity (sea clutter). The letter K refers to the modified Bessel function of the second kind. Physically, it comes again from an approximation of the sea surface by discrete scatterers. However, it allows for the population to fluctuate randomly, *i.e.* the number of discrete scatterers contributing to the scattered signal is itself a random variable (following a negative binomial distribution). This is known as the random walk model with fluctuating step number. The K distribution is then obtained by making the *average* number of scatterers go to infinity. Allowing for step number (population) fluctuations is a way to modulate the local average power of the sea clutter and accounting for longer timescale phenomena (swell for example). Note that the K distribution has been extensively validated by real data (again see [141] and references therein).

Even though its development was motivated physically by the presence of two time scales (the Gaussian clutter, and the local modulation of the power), it is a bit misleading to talk about time for the K distribution. Indeed, it is a *static* model, in the sense that it is valid for any fixed time t. If t is arbitrary and z_t denotes the squared modulus of the sea clutter at time t, then $z_t \sim K$ distribution. However, if one takes two different times t_1 and t_2, then the K distribution is unable to express the correlation between z_{t_1} and z_{t_2}, which could for example be represented by the joint distribution of z_{t_1} and z_{t_2}. The staticity may be the most notable limit of the K distribution.

A *dynamic* extension of the K distribution has been developped by T. R. Field [48], referred to as *Field's model* hereafter. It is also based on the random walk model, except that it models the dynamics of the scatterers (not only their one time distributions). Field's model is promising for two reasons: first, it is an extension to the K distribution and as such inherits all its advantages, and second it is dynamic and therefore relevant to model a dynamic phenomenon (sea clutter). In Field's model, the dynamics are expressed by *stochastic differential equations* (SDE). SDE are similar to ordinary differential equations but they include randomness *via* a volatility component. The solution of a SDE is a Markov stochastic process, often called a diffusion. It is possible to quickly get a grasp of what SDE are, for example by looking at numerical schemes to solve them. However, their rigorous definition is based on Itô integration theory of stochastic processes (see [105], [86], [54]) and requires notions which are relatively advanced compared to the usual notions of the signal processing community.

We can cite an example in which the motion of the sea surface has revealed problematic: Synthetic Aperture Radar imaging (SAR) [19], [97], [94]. In SAR, a radar is fixed on an airplane or a satellite and emits radar waves toward the ground or sea. Because the radar has a large aperture in the direction of flight, every point is revisited multiple times by the radar as the satellite or airplane passes by. SAR algorithms exploit this fact by integrating the measures made at different positions, and as a consequence dramatically improve the azimut resolution of the radar image. If this technique as proven useful on land, there is evidence that it is not as adequate over the sea because of the motion of the sea surface during the integration, causing a blurring effet [77], [78]. This example is a strong motivation for the following work, because it is directly related to the dynamicity of the sea surface and therefore of the sea clutter.

In this thesis, we propose to study Field's model from a remote sensing and signal processing perspective with a broad view, not limited to SAR. We aim at clarifying it by giving more details than [48] does, and work our way toward making it an applicable and profitable tool for sea clutter time series analysis. For example, we compute the probability of transition from z_{t_1} to z_{t_2} and explain that it can be used to infer z_{t_2} given that only z_{t_1} is observed. Taking into account a current observation to infer a future one was not possible with the K distribution because of its staticity. Generally speaking, we have an engineer approach, in the sense that mathematical rigour has sometimes to be set aside for making progress. Bearing this in mind, we explain how Field's model represents the sea clutter and derive results which are *necessary* to interpret and analyse data in the light of Field's model.

In the first chapter, we introduce basic electromagnetism theory and explain the general situation of radar remote sensing of the sea surface. We define classical vocabulary from the

radar world, and present sea clutter. In the last section, we detail the random walk model/K distribution.

The second chapter is a mathematical parenthesis, where we define all the notions necessary to understand stochastic differential equations, and more generally to understand the rest of the thesis. It may seem overburdened, but we think that, at times, precise definitions are helpful if not indispensable. Besides stochastic differential equations, we define a second important concept at the heart of our work: transition probabilities.

At this point, we know what sea clutter, the random walk model and stochastic differential equations are. We are in a place where it is possible to introduce Field's model, which is the purpose of the third chapter. We present it in a different, hopefully more pedagogic way than it is in [48]. We explain that in Field's model, the complex reflectivity Ψ_t (sea clutter) can be factorized as $\Psi_t = x_t^{1/2}\gamma_t$, where x_t is the radar cross section and γ_t the (complex) speckle. Both solve different SDE, parameterized by two parameters $\mathcal{A}$ and α for x_t and one parameter $\mathcal{B}$ for γ_t. The model has essentially two timescales carried respectively by x_t (slow local power modulation) and γ_t (fast phase decoherence), and which are retrieved in the fact that $\mathcal{A} \ll \mathcal{B}$.

In the fourth chapter, we introduce and solve the Fokker-Planck equations for x_t and γ_t, to get their transition probabilities. Using numerical simulations, we illustrate the progressing spreading (diffusion) of the transition probability starting from a dirac initial condition. We also compute the transition probabilities of other processes related to the sea clutter, such as $z_t = |\Psi_t|^2$. In this chapter, the transition probabilities are viewed as a mean to predict future (or past) values of the sea clutter given a present time observation. We argue that this opens the possibility of synchronizing measurements made at different positions and times, as this is the case in Synthetic Aperture Radar imaging.

In the fifth chapter, we derive maximum likelihood estimators of $\mathcal{A}$ and $\mathcal{B}$ from finite duration discrete time series of x_t and γ_t respectively, and ergodicity-based estimators for α. We use numerical simulations to assess and compare the performance of the estimators. The maximum likelihood estimation of $\mathcal{A}$ and $\mathcal{B}$ requires transition probabilities to compute the likelihood functions. Naturally, we use the exact transition probabilities obtained in the fourth chapter, but also compare the results to two approximations. In particular, we show numerically that using Gaussians to approximate the transition probabilities is completely satisfying in our context.

In the sixth chapter, we adapt Field's model to account for the presence of a target either with constant reflectivity (homodyned K scattering, HK), or with a reflectivity modulated by x_t (generalized K scattering, GK). In both cases, we derive SDE for the sea + target reflectivity, and use Euler-Maruyama's scheme to approximate the transition probabilities. We compute the likelihood of a time series and derive maximum likelihood estimators for the target parameters, which are tested and approved numerically. We show that the maximum likelihood approach to estimating the target parameter is especially valuable in GK scattering.

In the seventh chapter we address an issue which was disregarded so far: the non-observability of x_t, and therefore γ_t, since in practice a radar can observe Ψ_t but not x_t and γ_t separately. Using the timescale difference between x_t and γ_t, we derive an estimator for x_t based on observable data. To connect with the fifth chapter, a sequence of estimators for $\mathcal{A}$, $\mathcal{B}$, α (clutter parameters), x_t (radar cross section) and γ_t (speckle) is proposed and tested numerically. We show that bayesian estimation of the clutter parameters is straightforward and we explicit two methods based on Field's model for target detection.

Finally, we conclude and give directions for future research.

Chapter 1

Sea clutter

In this chapter, we define the sea clutter and present the random walk model. The term sea clutter is truly used for the first time in section 1.2.6 and refer to the radar signal scattered by the sea surface. Since radar waves are electromagnetic waves (EM waves), we first present some basic elements of electromagnetism in section 1.1. In section 1.2, we introduce the general context of radar remote sensing of the sea surface and define the *complex reflectivity*. In section 1.2.6, we see with real data that the sea surface complex reflectivity looks like noise at first sight, hence the term *sea clutter*. We present the Stratton-Chu equations for the scattered electromagnetic field. Because we have a signal processing point of view, we choose to discard them due to their relative complexity compared to other models. Once we have observed the stochastic nature of the sea surface reflectivity and renamed it sea clutter, we present the random walk model in section 1.3. It is the first building block toward a fully dynamic stochastic model as exposed in chapter 3 (Field's model).

1.1 Basics of Electromagnetism

In this section, we introduce elements of electromagnetic theory which are useful to understand the physical basis of radar remote sensing: Maxwell's equations, polarization for plane EM waves, and the scattering cross section for spherically decaying EM waves. Only monochromatic EM waves are mentionned. Refer to [60] for a detailed account of the basics of electromagnetism.

1.1.1 Maxwell's equations

Let $\vec{E}$ be the *electric field* and $\vec{B}$ the *magnetic induction*. Maxwell's equations form a coupled systems of differential equations for $\vec{E}$ and $\vec{B}$:

$$\nabla \cdot \vec{E} = \frac{\rho}{\epsilon_0}, \tag{1.1}$$

$$\nabla \cdot \vec{B} = 0, \tag{1.2}$$

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \tag{1.3}$$

$$\nabla \times \vec{B} = \mu_0 \left(\vec{J} + \varepsilon_0 \frac{\partial \vec{E}}{\partial t} \right). \tag{1.4}$$

∇ is the nabla operator, ρ is the density of electric charge (charge per unit volume), $\vec{J}$ is the density of electric current (current per unit area), ϵ_0 is the electric permittivity of free space, and μ_0 is the magnetic permeability of free space. c is the speed of light in free space ($c = 299\,792\,458$ m/s). There is a well-known relation:

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}. \tag{1.5}$$

In a material, it is convenient to define $\vec{D} = \epsilon_0 \epsilon \vec{E}$ the *electric displacement* ($\epsilon_0 \epsilon$ being the permittivity of the medium) and $\vec{H} = \frac{1}{\mu_0 \mu} \vec{B}$ the *magnetic field* ($\mu_0 \mu$ being the permittivity of the medium) . Then the "macroscopic" Maxwell's equations can be written:

$$\nabla \cdot \vec{D} = \rho, \tag{1.6}$$

$$\nabla \cdot \vec{B} = 0, \tag{1.7}$$

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \tag{1.8}$$

$$\nabla \times \vec{H} = \vec{J} + \frac{\partial \vec{D}}{\partial t}. \tag{1.9}$$

1.1.2 Polarization

A special solution of the Maxwell's equations is the *plane monochromatic wave* propagating in some direction $\vec{z}$. In the orthonormal basis $(\vec{x}, \vec{y}, \vec{z})$, the electric field is represented by $\begin{bmatrix} E_x & E_y & E_z \end{bmatrix}^\top$

$$\begin{cases} E_x(z,t) = E_{0_x} \cos(\omega t - kz + \phi_x) \\ E_y(z,t) = E_{0_y} \cos(\omega t - kz + \phi_y) \\ E_z(z,t) = 0. \end{cases} \tag{1.10}$$

In complex notation:

$$\begin{cases} E_x(z,t) = E_{0_x} e^{i(\omega t - kz + \phi_x)} \\ E_y(z,t) = E_{0_y} e^{i(\omega t - kz + \phi_y)}. \end{cases} \tag{1.11}$$

If ϕ_x is used as a reference phase, this can be further denoted:

$$\begin{bmatrix} E_x(z,t) \\ E_y(z,t) \end{bmatrix} = E_0 e^{i(\omega t - kz + \phi_x)} \begin{bmatrix} V_x \\ V_y \end{bmatrix}. \tag{1.12}$$

$\begin{bmatrix} V_x \\ V_y \end{bmatrix}$ is the normalized Jones vector. Table 1.1 gives well-known examples of Jones vector, with the corresponding polarization. We will mostly refer to linear polarization in this thesis. However, there is no loss of generality since it is always possible to decompose an elliptically polarized EM wave into a sum of two linearly polarized waves. Two directions for linear polarization are usually referred to in radar remote sensing (section 1.2): horizontal (H) and vertical (V). They are represented in figure 1.1.

If the EM wave propagate in the $\vec{n}$ direction ($\vec{n} = \vec{z}$ above), it can be shown that $\vec{E} = \begin{bmatrix} E_x & E_y & E_z \end{bmatrix}^\top$, $\vec{B} = \begin{bmatrix} B_x & B_y & B_z \end{bmatrix}^\top$ (complex representations) and the direction propagation $\vec{n}$ (for example $\vec{z}$ above) are related by:

$$\vec{B} = \frac{1}{c} \vec{n} \wedge \vec{E}. \tag{1.13}$$
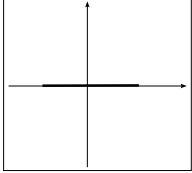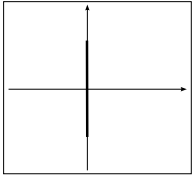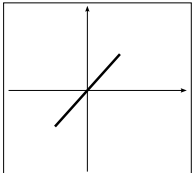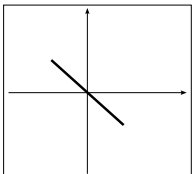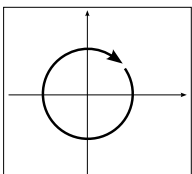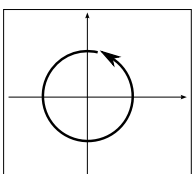
Polarization state	Normalized Jones vector	Representation
Horizontal linear	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$	
Vertical linear	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$	
Linear at $+45°$	$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$	
Linear at $-45°$	$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$	
Right circular	$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}$	
Left circular	$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}$	

Table 1.1: A few exemples of Jones vectors.
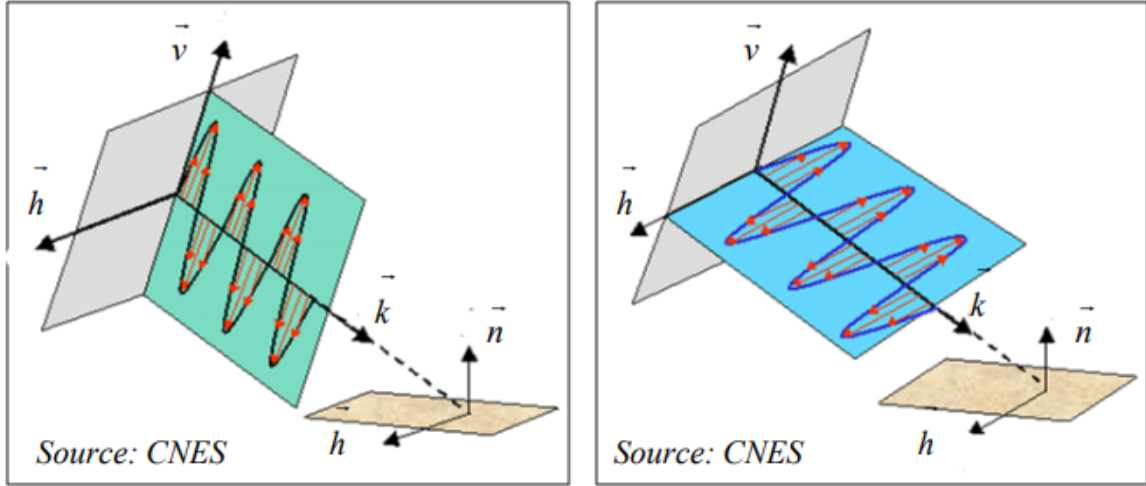
Figure 1.1: H (right) and V (left) polarized electromagnetic waves. The source is at the origin of the frame and the wave propagates in the $\vec{k}$ direction. Any elliptically polarized EM waves can be decomposed into a sum of an H and a V polarized EM wave.

It means in particular that $\vec{E}$ and $\vec{B}$ lie in the plane perpendicular to $\vec{n}$, such that only two coordinates are necessary to describe them. It also means that $\vec{B}$ is determined by $\vec{E}$ and $\vec{k}$, so we only need to care about $\vec{E}$.

1.1.3 Power and energy

One can compute the power transported by an EM wave in a vacuum by introducing the Poynting vector:

$$\vec{S} = \frac{\vec{E} \wedge \vec{B}}{\mu_0}. \tag{1.14}$$

$\vec{S}$ has the unit of Watts/m². It gives the instantaneous (oscillates at the time scale of the EM wave) power per unit surface transported by the EM wave, in the direction of propagation. For directions other than the direction of propagation, one shall multiply equation (1.14) by the cosine of the angle between the direction of propagation and the direction of computation of the flux. For a plane linearly polarized monochromatic wave propagating in free space, the time average (over one oscillation period) of the amplitude of the Poynting vector is:

$$\langle S \rangle = \frac{1}{2\eta_0} |E_m|^2, \tag{1.15}$$

where $\eta_0 = c\mu_0$ is the impedance of the transmission medium (vacuum here), and where the electric field is assumed to be:

$$\vec{E}(\vec{r}, t) = E_m \cos(\omega t - \vec{k}.\vec{r} - \phi)\vec{u}. \tag{1.16}$$

$\vec{u}$ is the polarization direction, and $\vec{k}$ is the wave vector, which points in the propagation direction. Equation (1.15) gives the transmitted power per unit surface area, in Watts/m².

We now want to define the notion of *scattering cross section*. To do so, we consider an emitter which emits spherical monochromatic EM waves. We assume that the emitter is at a

distance R from an object, and that R is large enough to allow us to locally approximate the EM waves by linearly polarized plane waves. The object is illuminated by the EM waves and scatters it. A part of the scattered waves are scattered back to the receiver. For simplicity, we consider that the receiver and emitter are at the same distance from the object. The total power received by an object of surface s oriented perpendicular to the wave propagation direction:

$$P_r = \frac{s}{2\eta_0 R^2}|E_m|^2. \tag{1.17}$$

The R^2 coefficient is for the geometric decay of the spherical wave, but locally, we use equation (1.15) for the average power carried by a plane wave. Figure 1.2 illustrates very well why the received power evolves as $\frac{1}{R^2}$. This attenuation, for spherical waves, is purely geometric and corresponds to the spreading of the constant total power (or intensity) over the surface of a sphere, which grows as R^2.
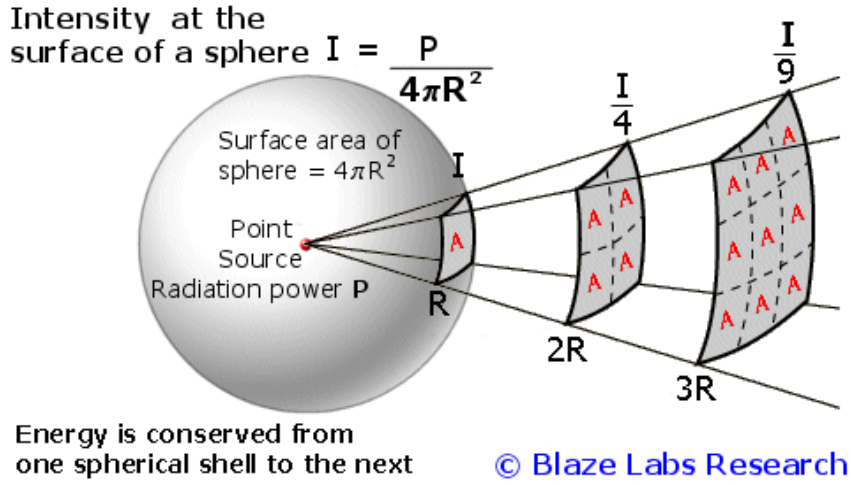
Figure 1.2: Illustration of the geometric spreading of the intensity of spherical waves.

If the object scattered all the power isotropically, then the backscattered power per unit surface would be:

$$\Pi_s = \frac{1}{4\pi R^2}P_r = \frac{1}{\eta_0 8\pi R^4}s|E_m|^2. \tag{1.18}$$

Then s can be retrieved by:

$$s = \frac{\eta_0 8\pi R^4 \Pi_s}{|E_m|^2}. \tag{1.19}$$

In equation (1.19), s is the surface of the object projected perpendicular to the direction of propagation of the incident waves. For objects which scatter the EM waves anisotropically, we cannot relate directly the surface of the object to the incident and scattered fields, but we can define the *scattering cross section* as:

$$\sigma = \frac{\eta_0 8\pi R^4 \Pi_s}{|E_m|^2}. \tag{1.20}$$

Because the object is anisotropic, Π_s and therefore σ depend on the scattering direction.

σ is the surface that the scattering object should have, if the received power had been scattered isotropically by it. We shall put equation (1.20) in a more familiar form. It is often

assumed that we start directly with the incident power per unit surface (without mentioning the emitted EM wave and its spherical spreading). In our case, the incident power per unit surface is:

$$\Pi_i = \frac{1}{2\eta_0} \frac{|E_m|^2}{R^2}. \tag{1.21}$$

In terms of Π_i, the scattering cross section can be written:

$$\sigma = \frac{4\pi R^2 \Pi_s}{\Pi_i}. \tag{1.22}$$

A radar system (introduced in the next section) emits electromagnetic waves with some linear polarization $\vec{u}$. Let denote $\vec{E}_{in}$ the incident electric field. The backscattered EM waves have arbitrary polarization, but the receiver projects the waves on some observation polarization $\vec{v}$. Let denote $\vec{E}_{re}$ the received electric field. The scattering cross section is then:

$$\sigma(\vec{u}, \vec{v}) = \frac{4\pi R^2 \langle \vec{E}_{re} . \vec{v} \rangle}{\langle \vec{E}_{in} . \vec{u} \rangle}. \tag{1.23}$$

For homogeneous surfacic objects, and in particular if they are large (such as the sea surface), it is useful to define the scattering cross section per unit surface of the object:

$$\sigma_{surf}(\vec{u}, \vec{v}) = \frac{\sigma(\vec{u}, \vec{v})}{A}, \tag{1.24}$$

where A is the area of the object.

The scattering cross section of an object depends on many parameters: the geometry of acquisition (see also figure 1.7), the polarization, the frequency, the shape and nature of the object etc.

1.2 Radar remote sensing of the sea surface

In this section, we present qualitatively the main situation of interest in this thesis: radar remote sensing of the sea surface. We expose the most important features of radar systems *from our time series signal processing point of view.* The general situation is represented in figure 1.3: a radar emits EM waves towards its surrounding environment, which scatters part of the EM waves to a receiver. The scattered *signal* contains information about the environment. The challenge is to process the signal correctly to extract that information. It is an act of remote sensing of the environment, which is usually done from the coast, a plane, or a satellite when the sensed environment is the sea surface (figure 1.3).

1.2.1 Radar acquisition chain

One can distinguish two types of radar configurations: monostatic and bistatic radars [131]. Only the monostatic configuration appeared in figure 1.3. In both cases, there is an emitter, which emits electromagnetic waves (EM waves) toward the object under study, and a receiver, which can receive the direct propagation of the emitted signal (which is of limited interest), and the EM waves scattered by the object. This second contribution is what we are interested
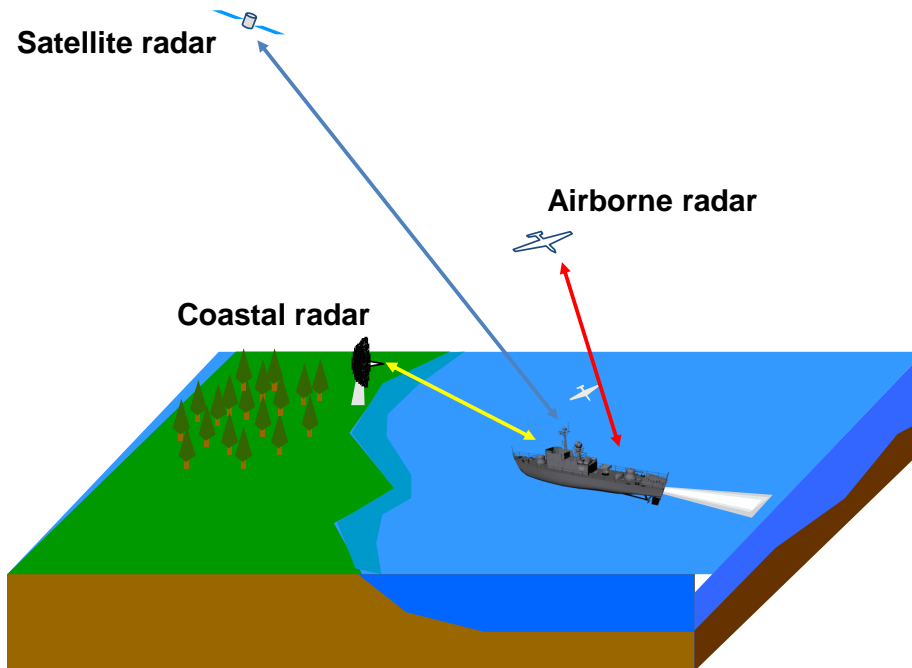
Figure 1.3: Coastal, airborne, and satellite radars. Extracted from [16].

in. In the monostatic configuration, the emitter and receiver are at the same position, while they are at different positions for the bistatic configuration. Figure 1.4 represents the different components of a bistatic radar acquisition chain in a maritime environment, where the scattering 'objects' can be the sea and/or a boat (possibly a whale, and emerged submarine etc. This is left to the reader's imagination). We will consider that the final product of this chain is the received signal.

Considerable work has been, and can still be done on each of the yellow and green boxes. Many electromagnetic models are available for relating the transmitted signal (upper left red box) to the received one. In this thesis, we are interested in one of them: Field's model, which is presented later in chapter 3. It constitutes our reference model all along our work. It can be understood as a black box relating directly the transmitted signal to the received one, with only a few parameters. By analogy, Field's model has three parameters for the lower left green box (sea parameters), which will be denoted $\mathcal{A}, \mathcal{B}$ and α. It has 2 parameters for the lower right green box of the ship characteristics (see chapter 6): $\Psi_c^{(R)}$ and $\Psi_c^{(I)}$. It does not have any parameters for the atmosphere (upper right green box).

We will clarify the relation between the transmitted and received signals through Field's model, and thanks to its simplicity, we will show that it leads to developments in the post-processing phase (blue box). We will often simulate numerically what can be interpreted as the received signal, $i.e.$ the output of the acquisition chain.
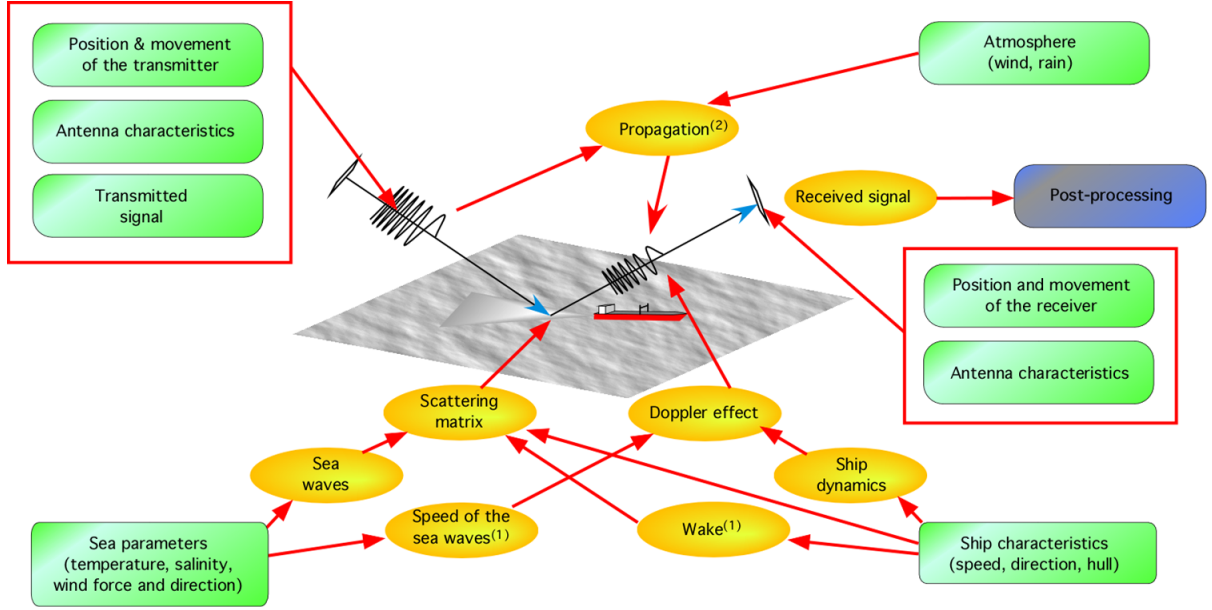
Figure 1.4: Detailed description of the radar acquisition chain. Extracted from [16].

1.2.2   Geometry of acquisition

The basic geometry of acquisition of a monostatic radar is represented in figure 1.5. The antenna (upper left) is rectangular and emits EM waves in all directions but anisotropically. In the direction given by the angles $\breve{y}_x$ and $\breve{z}_x$, the antenna gain is ([16] p 27):

$$\sigma_x(\breve{y}_x, \breve{z}_x) = S_X \text{sinc}\left(\pi\frac{d_z}{\lambda_0}k_y\right)\text{sinc}\left(\pi\frac{d_y}{\lambda_0}k_z\right).$$ (1.25)

d_y and d_z are the dimensions of the antenna in the $\hat{\mathbf{y}}_x$ and $\hat{\mathbf{z}}_x$ directions, $S_x = d_y d_z$ is the area of the antenna, λ_0 is the carrier wavelength and k_y and k_z are given by:

$$\begin{aligned} k_y &= \cos(\breve{y}_x)\sin(\breve{z}_x) \\ k_z &= \sin(\breve{y}_x). \end{aligned}$$ (1.26)

Equation (1.25) defines the *radiation pattern*. Though in theory this equation gives a non zero power in all directions, in practice we consider that only the region for which the gain relative to the maximum is greater than -3 dB, equivalent to a factor 0.707. It defines the ellipse-like region in figure 1.5, called the *footprint*. The size of the footprint is Δd_r in the $\hat{x}$ direction and Δd_{az} in the $\hat{y}$ direction. Δd_r and Δd_{az} are respectively the *ground range resolution* and *azimutal resolution*. In figure 1.5, θ_i is the *incidence angle* and $\frac{\pi}{2} - \theta_i$ is the *grazing angle*. It will be noticed in section 1.2.6 that the characteristics of the EM waves scattered by the sea surface depend on the grazing angle.

1.2.3   The emitted signal

We have seen in the previous section that the EM waves emitted by a radar is spatially spread according to some radiation pattern. In a solid angle where the radiation pattern is almost constant, the power of the EM waves decays geometrically as $\frac{1}{R^2}$, like spherical waves. If a large R is fixed, the EM waves received by an object have decayed like a spherical wave,
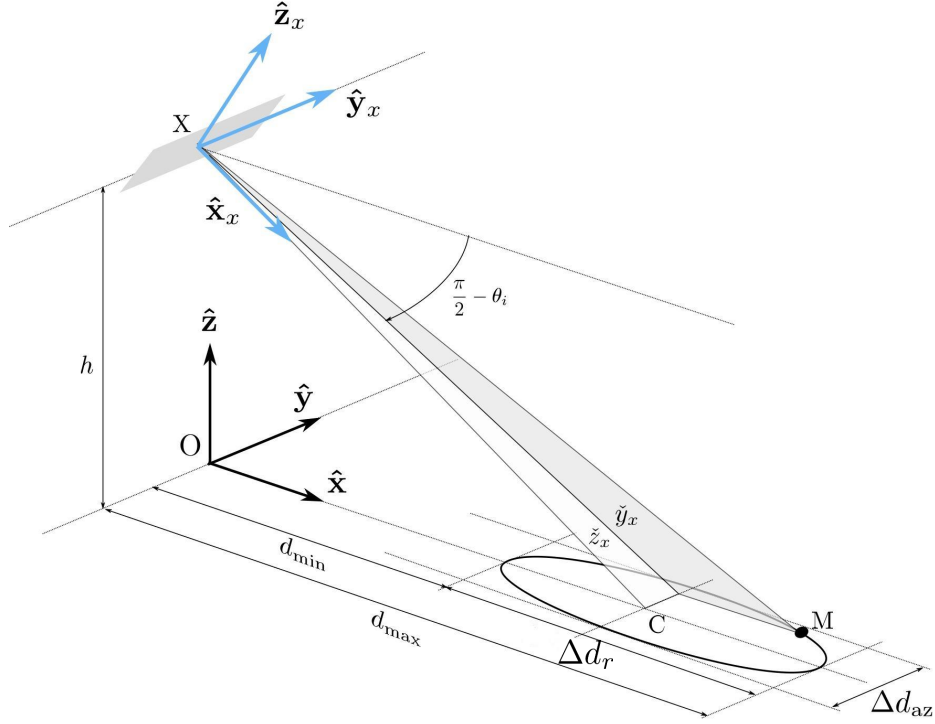
Figure 1.5: Geometry of radar acquisition featuring the antenna footprint and azimutal and ground range resolution. Extracted from [16].

but are locally plane and linearly polarized (section 1.1.2). In the direction of polarization, the electric field is simply $E(t)$ (one dimensional). At emission, the polarization is usually either H or V. $E(t)$ is not stricly monochromatic (*i.e.* a sinusoid). However, it is usually narrowband, *i.e.* its frequency content is centered around a carrier frequency f_0 and has a spread Δf such that $\Delta f \ll f_0$. Norms exist, which define frequency bands for radar waves. A commonly used norm is the Institute of Electrical and Electronics Engineers (IEEE) norm, depicted in table 1.2. Various wave forms are typically used: truncated sinus, chirp, frequency-modulated continuous-wave (FMCW) etc. A *pulse* is one short group of waves, with a sinus or chirp (or else) waveform. For example, the truncated sinus reads (in complex notations):

$$E_X^{sin, f_0}(t) = E_0 e^{j2\pi f_0 t} \mathbf{1}_{\left[-\frac{T_X}{2}, \frac{T_X}{2}\right]}(t), \tag{1.27}$$

where $\mathbf{1}$ is the indicator function, and T_X is the pulse duration. The chirp reads:

$$E_X^{chirp, f_0, \Delta f}(t) = E_0 e^{j2\pi\left(f_0 + \frac{\Delta f}{2T_X}t\right)t} \mathbf{1}_{\left[-\frac{T_X}{2}, \frac{T_X}{2}\right]}(t). \tag{1.28}$$

The instantaneous frequency of the chirp is the derivative of its instantaneous phase (over 2π), *i.e.*:

$$f(t) = f_0 + \frac{\Delta f}{T_X}t. \tag{1.29}$$

Therefore, the chirp in equation (1.28) has linearly increasing instantaneous frequency. Figure 1.6 represents a sinus waveform of duration 1 s, with frequency 20 Hz, and a chirp with instantaneous frequency 20 Hz at $t = 0$ and $\Delta f = 50$ Hz. The numerical values are invented for illustrative purposes. In particular, the relative change of frequency over the duration of

Standard Radar Frequency Nomenclature (IEEE Standard 521-2002)		
Band designator	Frequency	Wavelength
HF	3-30 MHz	10-100 m
VHF	30-300 MHz	1-10 m
UHF	300-1000 MHz	0.3 - 1 m
L	1-2 GHz	15-30 cm
S	2-4 GHz	7.5-15 cm
C	4-8 GHz	3.75-7.5 cm
X	8-12 GHz	2.5-3.75 cm
Ku	12-18 GHz	1.67-2.5 cm
K	18-27 GHz	1.11-1.67 cm
Ka	27-40 GHz	0.75-1.11 cm

Table 1.2: Standard radar bands as defined by IEEE.

the chirp is much lower in real application. For example, for satellite remote sensing, [35] quotes a value of 5.3 GHz for f_0, a frequency modulation rate of 0.5 MHz/μs, and a pulse duration of 40 μs. The resulting Δf is:

$$\Delta f = 0.5.10^6 \times 40.10^{-6} = 20 \text{ MHz}, \tag{1.30}$$

which is only a small fraction of the carrier $f_0 = 5.3$ GHz. The advantage of using a chirp, is that for equal T_X, it can lead to much better range resolution that the sinus waveform by use of matched filtering (see [35]). The number of pulses per second that a radar emits is the *pulse repetition frequency* (PRF). The order of magnitude of the pulse duration T_X is tens of μs (10^{-6} s), while the order of magnitude of the PRF is 1000 Hz for satellite and airborne applications (see [35]), *i.e.* one thousand pulses per second.

When the phase of subsequent pulses is controlled, the radar is said to be *coherent*. Otherwise, it is *incoherent*.

1.2.4 The received signal

The complex reflectivity

Let us denote $E_X(t)$ the emitted signal (a pulse, or a sequence of pulses for example), which corresponds to the component of the electric field along its direction of polarization. At reception, we choose an observation polarization direction, usually either H or V. Since there are two options at emission (H or V) and two at reception (H or V), there are four channels: HH, VV, VH, HV, where the first letter is for emission and the second for reception. Under the narrowband approximation, it is assumed that the received signal is in the form:

$$E_R(t) = \Psi(t)E_X(t - \tau). \tag{1.31}$$
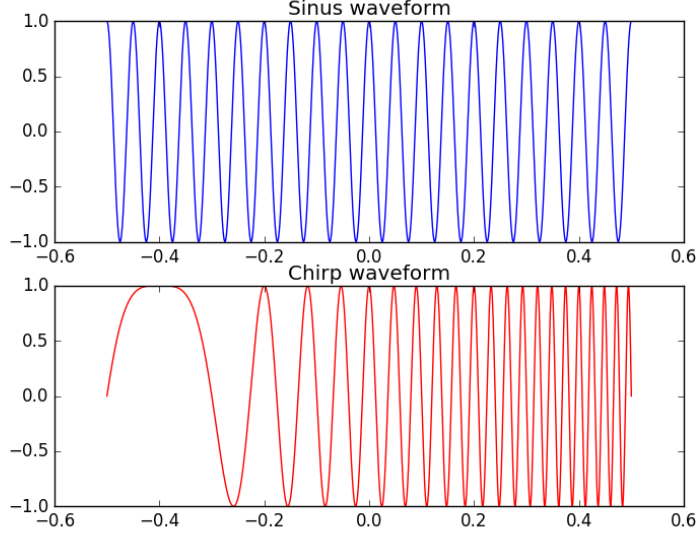
Figure 1.6: Comparison between the sinus (up) and chirp (down) waveforms.

$\Psi(t)$ is a complex coefficient which accounts for phase and amplitude changes. It is called the *complex reflectivity*. The reflectivity depends on the chosen emission and reception polarizations. Therefore, it should ideally be written with a subscript or a superscript to indicate which channel it is for: $\Psi(t)^{(HH)}$, $\Psi(t)^{(HV)}$ etc. τ is the delay. It is a fundamental assumption of this thesis that the scattered signal is of the form of equation (1.31). $\Psi(t)$ evolves much slower than $E_X(t)$, which, insides a pulse, oscillates at the frequency of the carrier. Physically, $\Psi(t)$ corresponds to the slow (compared to the carrier wave) variation of the scattering object reflectivity. It remains approximately constant over a pulse duration. If we consider one chirp centered at time t_c, we have:

$$E_X(t) = E_X^{chirp, f_0, \Delta f}(t - t_c), \tag{1.32}$$

and

$$E_R(t) = \Psi(t_c) E_X^{chirp, f_0, \Delta f}(t - t_c - \tau). \tag{1.33}$$

Let us consider a train of $n + 1$ chirps where the k-th chirp is centered at time t_k:

$$E_X(t) = \sum_{k=0}^{n} E_X^{chirp, f_0, \Delta f}(t - t_k). \tag{1.34}$$

Under the assumption that the delay of the k-th pulse is τ_k, the scattered signal is:

$$E_R(t) = \sum_{k=1}^{n} \Psi(t_k) E_X^{chirp, f_0, \Delta f}(t - t_k - \tau_k). \tag{1.35}$$

Equation (1.35) is even a strongest statement than equation (1.31), but we adopt it as well. It says that the chirps are all delayed by the some constant, or average τ, and that each of them is multiplied by some complex coefficient $\Psi(t_k)$. We write $\Psi(t_k) = \Psi_{t_k}$ to synchronize with the notations adopted from chapter 3 on. The discrete time series $\{\Psi_{t_k}, k = 0, 1, \ldots, n\}$ is the object of study of our thesis. In fact, Ψ_{t_k} is retrieved only if the radar is coherent

(pulse to pulse phase is controlled). If the radar is incoherent, we can retrieve only $|\Psi_{t_k}|$. The need for an adequate statistical model of Ψ_t when it is for the complex reflectivity of the sea surface is further explained in section 1.2.6.

What is the average modulus of Ψ_t? We understand that the average modulus of Ψ_t is much lower than one, *i.e.* the scattered wave has much weaker amplitude than the emitted wave. This is due to the geometric spreading of the energy described in section 1.1.3, the atmospheric attenuation, and the scattering characteristics of the object (the sea surface for us, and possibly a boat). In section 1.1.3, we defined the concept of scattering cross section for monochromatic waves, which we now call the *radar cross section* (RCS). The RCS is proportional to the average scattered power (equation (1.22)). It is important to mention that it is dependent to both the incident and observation directions. In figure 1.7, the incident EM wave propagates locally (close to the scattering object) in the direction $\hat{r}_x$. It is polarized linearly in the plane $(\hat{h}_i, \hat{v}_i)$. It is then scattered in all directions, but we only observe the field scattered around the direction $-\hat{r}_r$. The RCS of the scattering object depends on $\hat{r}_x$ and $\hat{r}_r$, the illumination and observation directions $\sigma = \sigma(\hat{r}_x, \hat{r}_r)$. The average value of Ψ_t is proportional to the RCS:

$$\langle \Psi_t \rangle \propto \sigma(\hat{r}_x, \hat{r}_r) \propto P_s(\hat{r}_x, \hat{r}_r), \tag{1.36}$$

where $P_s(\hat{r}_x, \hat{r}_r)$ is the average scattered power. It is possible to define a RCS which is dynamic, as we will see in subsequent chapters. It all depends on the length of the averaging window in equation (1.36), *i.e.* on what timescale we study Ψ_t.
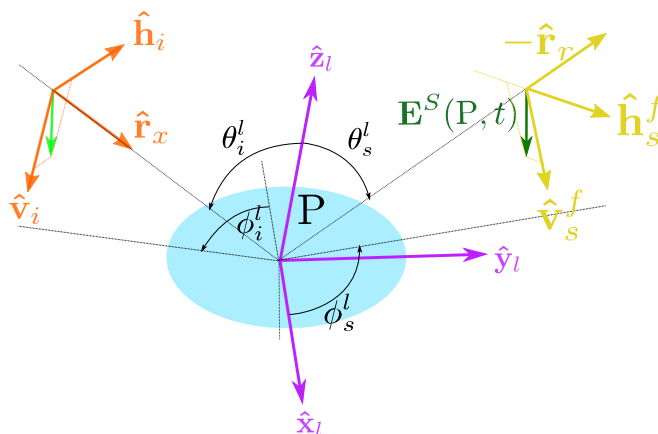
Figure 1.7: Representation of the frames used to express the incident and scattered EM waves [16].

The Doppler effect

If the scattering object and radar have a relative motion, the scattered waves may be subject to the famous *Doppler effect*. Due to the low velocities (relative to the speed of light) involved in radar remote sensing of the sea surface, we can neglect relativistic effects and make the galilean approximation.

Assume that the radar and the object have a relative motion given by the velocity vector $\vec{V}$ and that their relative position vector is $\vec{r} = -r(t)\vec{u}_r$ ($\vec{u}_r$ points from the object to the

radar). We have:

$$\frac{\mathrm{d}r(t)}{\mathrm{d}t} = \vec{V}.\vec{u}_r. \tag{1.37}$$

For a monostatic radar, the time derivative of the delay τ is:

$$\frac{\mathrm{d}\tau}{\mathrm{d}t} = 2\frac{\vec{V}.\vec{u}_r}{c_0}, \tag{1.38}$$

where the 2 factor is due to account for the round trip of the EM waves. If a element of wave is emitted at time t and another one at time $t + \delta t$, then the difference in their travel times is:

$$\delta\tau \approx \frac{\mathrm{d}\tau}{\mathrm{d}t}\delta t = 2\frac{\vec{V}.\vec{u}_r}{c_0}\delta t. \tag{1.39}$$

In fact, we could have postulated equation (1.39), which is more intuitive than equation (1.38). Equation (1.38) should be proven by first deriving an expression for τ, and then differentiating it. Now to get a better grasp of what the relative motion implies, we assume that a sinus pulse of duration T_c and carrier frequency f_0 is emitted toward the object. The difference in the travel time of both extremities of the pulse is approximately:

$$\frac{\mathrm{d}\tau}{\mathrm{d}t}T_c, \tag{1.40}$$

so the pulse duration at reception is:

$$T_R = T_c + 2\frac{\vec{V}.\vec{u}_r}{c_0}T_c = \rho_{Doppler}T_c, \tag{1.41}$$

with $\rho_{Doppler} = 1 + 2\frac{\vec{V}.\vec{u}_r}{c_0}$. Therefore, if $\vec{V}.\vec{u}_r$ is negative (radar and object getting closer), the pulse is compressed. We now look at the frequency of the oscillations within the pulse. At emission, two successive peaks are separated by the time interval $T_0 = 1/f_0$. At reception, they are received with a delay difference of:

$$\frac{\mathrm{d}\tau}{\mathrm{d}t}T_0, \tag{1.42}$$

such that the apparent period at reception is:

$$T_D = T_0 + 2\frac{\vec{V}.\vec{u}_r}{c_0}T_0. \tag{1.43}$$

The apparent frequency at reception is therefore:

$$f_D = \frac{1}{T_0 + 2\frac{\vec{V}.\vec{u}_r}{c_0}T_0} = \frac{f_0}{1 + 2\frac{\vec{V}.\vec{u}_r}{c_0}} = f_0\frac{c_0}{c_0 + 2\vec{V}.\vec{u}_r}. \tag{1.44}$$

If the radar and the object are getting closer, the denominator is reduced and the frequency f_D is larger than f_0 (and inversely if they are drifting apart). Figure 1.8 illustrates these two components of the Doppler effect: pulse compression (or dilatation) and frequency shift.

We will see in section 1.3 that it is relevant to approximate the sea surface by a population of discrete scatterers. In the case of airborne or satellite radar remote sensing, the uniform motion of the aircraft or satellite relative to the scatterers implies a large Doppler shift of the scattered wave frequency, while for a coastal radar, this shift (due to the radar motion) is zero. The motion of the scatterers themselves also come into play and explains that the scattered wave spectrum spreads around its mean value.
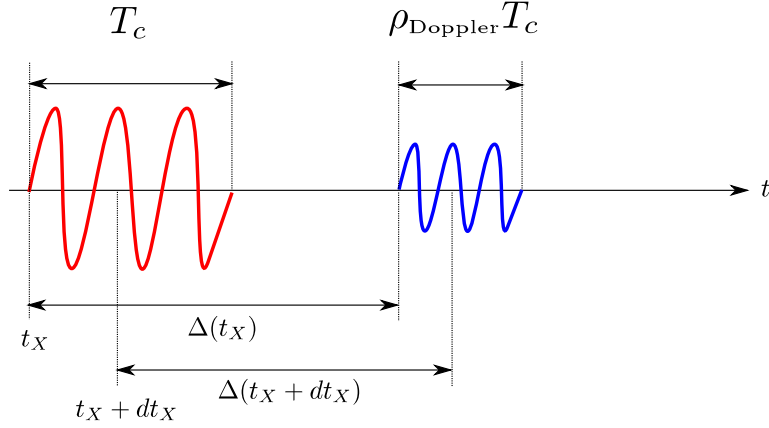
Figure 1.8: Illustration of the Doppler effect on a pulse. The emitted pulse is in red and the received one in blue.

1.2.5 Synthetic Aperture Radar imaging

Now that we are a bit accustomed with the basic radar theory, we can move to the subject of synthetic aperture radar imaging (SAR imaging). Our presentation is extremely concise. We aim at giving an idea of how it works, essentially because it could be one possible application of the work developped in this thesis (see chapter 4). For readers interested in the theory of SAR imaging, see [94], [97], [19]. Uncountably many other books exist on the topic and the reader could refer to any of them.

Synthetic Aperture Radar (SAR) imaging consists in creating a radar image using algorithms to dramatically improve the azimut resolution. It can be used for airborne and satellite radar remote sensing, *i.e.* when the platform is moving relative to the scene. In section 1.2.2 and figure 1.5, we saw that the antenna emits radar waves mainly in preferred directions (due to the radiation pattern) which makes a footprint of size Δd_r in ground range and Δd_{az} in azimut.

Improving the range resolution (slant range or ground range) is immediate: T_X being the duration of a pulse, any two objects separated in ground range by more than $\frac{T_X}{2c_0 sin\theta_i}$ can be distinguished (we remind that θ_i is the incidence angle). It is possible to do better than that using matched filtering, if for example the pulse is a chirp. However, it is more complicated to improve the azimut resolution Δd_{az}. We can do it though using SAR, which makes use of the fact that due to the width Δd_{az} of the radiated beam, a point P of the environment is illuminated for some time, and receives and scatters back several subsequent pulses. Roughly speaking, integrating theses pulses can lead to a dramatic compression of Δd_{az} into a new azimut resolution dr_{az}. dr_{az} is obtained synthetically (integration of the pulses) and corresponds theoretically to the azimut resolution of a radar with a much narrower aperture than the true radar. We can understand why the name *synthetic aperture radar*. The geometry of acquisition for an airborne radar is represented in figure 1.9. The radar radiates to the lower right relative to the direction of flight $\hat{y}$.

For the sake of the presentation, we go back and write a pulse as:

$$E_X(t) = \mu(t)e^{j2\pi f_0 t}, \tag{1.45}$$

in order to put the carrier wave $e^{j2\pi f_0 t}$ in factor and let the pulse have an arbitrary waveform. In *baseband*, it is simply:

$$E_X(t) = \mu(t), \tag{1.46}$$

where we keep the same notation E_X for simplicity. It is immediate that equations (1.27) and (1.28) can be rewritten in the form of equation (1.45). The received signal is:

$$E_R(t) = \Psi\mu(t-\tau)e^{j2\pi f_0(t-\tau)} = \Psi\mu(t-\tau)e^{j2\pi f_0 t}e^{-j2\pi f_0\tau} = E_X(t-\tau)e^{-j2\pi f_0\tau}, \tag{1.47}$$

or in baseband:

$$E_R(t) = \Psi\mu(t-\tau)e^{-j2\pi f_0\tau} = E_X(t-\tau)e^{-j2\pi f_0\tau}. \tag{1.48}$$

Ψ is the complex reflectivity of the object taken at a relevant time.
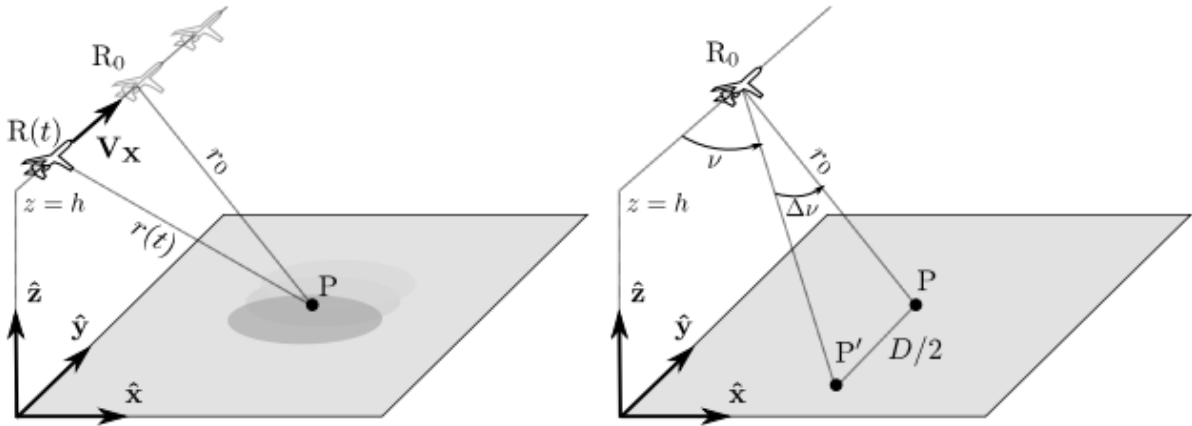
Figure 1.9: Geometry of a SAR acquisition.

Let $P = (x_P, y_P, 0)$ be an observed point, and $(0, V_X t, h)$ the coordinates of the radar as a function of t. Under the locally flat Earth approximation, the distance between the radar and P is:

$$
\begin{aligned}
r(t) &= \sqrt{x_P^2 + (y_P - V_X t)^2 + h^2} \\
&= \sqrt{r_0^2 + (y_P - V_X t)^2}, \tag{1.49}
\end{aligned}
$$

with $r_0^2 = x_P^2 + h_0^2$, the *distance of closest approach* (figure 1.9). If a pulse (number k) is emitted at time t_k, from equation (1.48) the received signal in basedband is:

$$E_R(t_k + t + \tau(t_k)) = \Psi_{t_k} E_X(t_k + t)e^{-j2\pi f_0\tau(t_k)}. \tag{1.50}$$

We take the reflectivity at time t_k but it could be taken at time $t_k + \tau(t_k)/2$ if $\tau(t_k)$ is not negligible relative to the time scale of the reflectivity. The delay now depends on t_k and we have:

$$\tau(t_k) = \frac{2r(t_k)}{c_0}, \tag{1.51}$$

such that:

$$E_R(t_k + t + \tau(t_k)) = \Psi_{t_k} E_X(t_k + t)e^{-j\frac{4\pi}{\lambda_0}r(t_k)}. \tag{1.52}$$

We assume that $y_P = 0$, in which case r_0 is reached at $t = 0$. We have $r(t) = \sqrt{r_0^2 + V_X^2 t^2}$. The second order expansion of $r(t_k)$ gives:

$$r(t_k) \approx r(0) + \frac{\partial r}{\partial t}(0)t_k + \frac{1}{2}\frac{\partial^2 r}{\partial t^2}(0)t_k^2. \tag{1.53}$$

Since we have:

$$\frac{\partial r}{\partial t}(t) = \frac{V_X^2 t}{\sqrt{V_X^2 t^2 + r_0^2}} = \frac{V_X^2 t}{r(t)} \tag{1.54}$$

$$\frac{\partial^2 r}{\partial t^2}(t) = \frac{V_X^2 r(t) - \frac{\partial r}{\partial t}(t)V_X^2 t}{r(t)^2}, \tag{1.55}$$

equation (1.53) becomes:

$$r(t_k) \approx r_0 + \frac{1}{2}\frac{V_X^2 t_k^2}{r_0}. \tag{1.56}$$

Injecting (1.56) into (1.52), we get:

$$E_R(t_k + t + \tau(t_k)) = \Psi_{t_k} E_X(t_k + t)e^{-j\frac{2\pi}{\lambda_0}\frac{V_X^2 t_X^2}{r_0} - j\frac{4\pi}{\lambda_0}r_0}. \tag{1.57}$$

We will not go further in the theory in this short introduction to SAR. What matters is to notice that we arrive at an expression for the *phase history* contained in the exponential of equation (1.57). SAR algorithms assume a constant reflectivity Ψ_{t_k} and make use of this phase history to obtain an accurate azimut resolution (indeed we obtain a chirp signal in the azimut direction to which we can apply matched filtering). Doing so, if the antenna length is L, the achievable azimut resolution is:

$$\Delta dr_{az} = \frac{L}{2}. \tag{1.58}$$

For satellite radars, Δd_{az} is in order of kilometers, while Δdr_{az} is in order of meters. Notice that for simplicity we have completely ignored the Doppler effect.

There are different algorithm for reconstructing SAR images. A famous SAR algorithm is the Range Doppler Algorithm (see [35] chapter 6). Figure 1.10 is an example of a SAR image, here of the Strait of Gibraltar. The interpretation of SAR images can be delicate, despite its apparent ressemblance to optical images. That being said, we can clearly distinguish Spain, Morocco, and the ocean/sea on figure 1.10.

1.2.6 Phenomenology of sea clutter

In the previous section, we presented SAR imaging for airborne or satellite radar. The cross track partition into resolution cells was not discussed but we said that it was immediate to obtain it. The along track partition into resolution cells involves SAR integration algorithms. Obtaining a radar image involves many steps, such that the image cannot really be considered as raw data. Let us now focus on the sea surface only. If a radar illuminates a fixed portion of the sea with some fixed Pulse Repetition Frequency (PRF), what we get from the backscattered signal is a discrete time series $\{\Psi_{t_k}, k = 0, 1, \ldots, n\}$ of the reflectivity (for coherent radars, $\{|\Psi_{t_k}|, k = 0, 1, \ldots, n\}$ otherwise) where $t_k - t_{k-1} = \frac{1}{PRF}$ for all k. It corresponds to the seemingly static situation of figure 1.5.

Figure 1.10: SAR image of the Strait of Gibraltar (ESA).

We think that the most instructive way to understand the particularities of radar remote sensing *of the sea surface* is to plot actual time series. Figure 1.11 and 1.12 represent time series of the complex reflectivity recorded by Ifremer C-band radar (we are grateful to Louis Marié, Bertrand Chapron and Frédéric Nougier for the data). The radar emits coherent chirps at carrier frequency $f_0 = 5.3$ GHz, and the PRF is 200 Hz. 4 polarizations are recorded simultaneously by the radar: HH, VV, HV, VH (see section 1.1.2). The first letter is for the emission polarization, and the second for reception. Figure 1.11 represents the real and imaginary parts of the HH (normalized) reflectivity for a duration of 3 seconds. The most striking feature is that it looks like noise. That is why the reflectivity of the sea surface is from now called *sea clutter*. Of course, seeing it like noise may be mathematically convenient, but in reality the sea surface has structures at all time and length scales, which evolve, interact, exchange energy etc. The swell is an example of structured motion which cannot be considered as noise but is deterministic. Whether a signal should be considered as noise is an interesting question and depends on its complexity and on the scale of observation. The real and imaginary parts of the sea clutter are not obviously correlated from figure 1.11. Because the complex reflectivity is observed (with its real and imaginary parts, or modulus and phase), figure 1.11 is representative of what a coherent radar may record. Figure 1.12 represents the modulus of the complex reflectivity for the four channels, which is representative of what an incoherent radar could measure. There is no obvious correlation between different channels.

Sea clutter depends on many parameters. Amongst others, it depends on the carrier
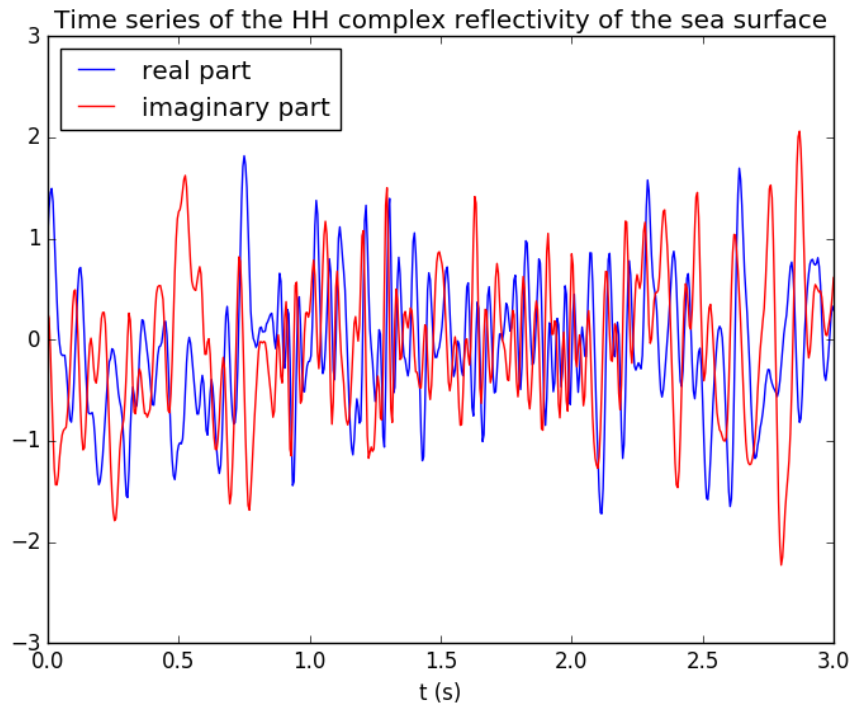
Figure 1.11: Time series of the complex reflectivity of the sea surface recorded by Ifremer C-band radar in the HH channel.
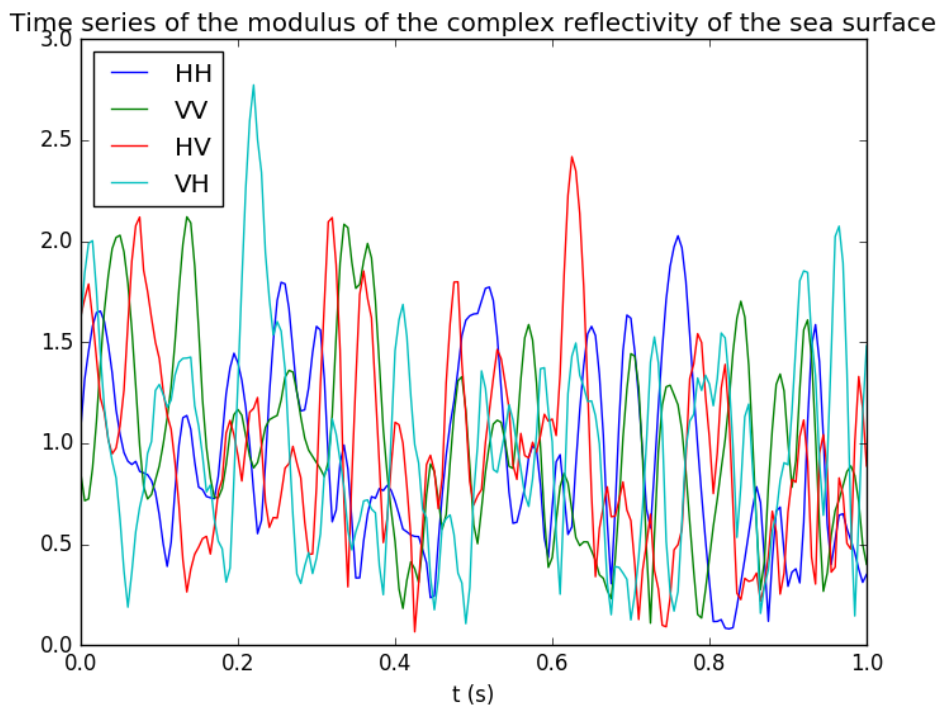
Figure 1.12: Time series of the modulus of the complex reflectivity of the sea surface recorded by Ifremer C-band radar in four different channels: HH, VV, HV, VH.

frequency f_0, the emission and reception polarizations, and the geometry of acquisition (orientation and grazing angle). Qualitatively, the sea surface is a superposition of many structures from the centimeter level (capillary waves) to the kilometer level, as evidenced by the large scale waves visible in the Mediterranean sea in figure 1.10. The point of view that we adopt is to consider that the EM waves are mainly scattered by discrete scatterers, which is the point of view adopted in [141] and in the random walk model presented in the next section. We will not precise what the scatterers are, but we need to remember that the population of scatterers are a priori different for different channels, f_0 and orientation and grazing angles. If the footprint of the radar is large enough, many scatterers will contribute to the reflectivity. Since the sea surface is dynamic, the scatterers move, appear, disappear, leading to an unpredictable sea clutter. For electromagnetic scattering by discrete scatterers, see [70], [48] and [99].

1.2.7   The Stratton-Chu equations

In the previous section, we have seen with real data that the sea surface reflectivity looks unpredictable and gives noise-like time series, hence the term clutter. In this section, we see that we can derive a theoretical expression of the field scattered by the sea surface in terms of an integral over the surface. We encourage the reader to refer to [142] chapter 17 in priority, and to [136] and [113].

Scalar scattering

It is easier to present the problem for scalar scattering and then admit the equations for vector scattering. Let $S(t)$ denote the sea surface, its height being $h = h(x, y, t)$. Consider a problem which is independent of one of the three spatial dimensions. For example, the surface may be corrugated such that $h = h(x, t)$, and the sources may be infinite in the y direction. In that case, the scattering problem is 2D, which implies that if the electric field is monochromatic and horizontal, we have in baseband:

$$\vec{E} = E(x, z)\vec{y} = E(x, z)\vec{h}, \tag{1.59}$$

$\vec{h}$ being a unitary vector defined in figure 1.2. Let $\vec{r}$ be the position vector. It can then be proven that E is solution of the Helmoltz equations:

$$\nabla^2 E(\vec{r}) + k_1^2 E(\vec{r}) = \xi(\vec{r}) \quad above\ the\ surface \tag{1.60}$$

$$\nabla^2 E(\vec{r}) + k_2^2 E(\vec{r}) = 0 \qquad below\ the\ surface, \tag{1.61}$$

where $\xi(\vec{r})$ is the source and $k_1^2 = \frac{\omega^2}{c^2}$, $k_2^2 = \frac{\omega^2}{c_s^2}$ with c_s the speed of light in the sea. Green's theorem, which can be derived itself from Gauss's theorem, states that for two scalar functions ϕ and ψ, and for a surface S enclosing a volume V, it holds:

$$\oint_S \left(\phi(\vec{r})\nabla\psi(\vec{r}) - \psi(\vec{r})\nabla\phi(\vec{r})\right).\mathrm{d}\vec{S} = \int_V \left(\phi(\vec{r})\nabla^2\psi(\vec{r}) - \psi(\vec{r})\nabla^2\phi(\vec{r})\right)\mathrm{d}V. \tag{1.62}$$

We remind that by definition, Green's function $G_0(\vec{x}, \vec{x}')$ solves:

$$\nabla^2 G_0(\vec{x}, \vec{x}') + k_1^2 G_0(\vec{x}, \vec{x}') = \delta(\vec{x} - \vec{x}'), \tag{1.63}$$

where δ is the Dirac distribution. We denote $E(\vec{r})$ the total field (scattered + incident) at $\vec{r}$ (position of the receiver), and E_{in} the incident field. It can be shown, using Green's theorem and the definition of Green's function that:

$$E(\vec{r}) = E_{in}(\vec{r}) + \int_{S(t)} \left(G_0(\vec{r}, \vec{x}')\nabla' E(\vec{x}').\vec{n}(\vec{x}') - E(\vec{x}')\vec{n}(\vec{x}').\nabla' G_0(\vec{r}, \vec{x}')\right) \mathrm{d}S(\vec{x}'). \quad (1.64)$$

We insist that $S(t)$ is the sea surface over which we integrate to compute the scattered field in equation (1.64). It is different from the surface S (enclosing the volume V) of equation (1.62). $\vec{n}(\vec{x}')$ is the normal to the surface at $\vec{x}'$ (point on the surface). We write ∇' to specify that the gradient is taken relative to the primed variable. Equation (1.64) is not explicit, since the gradient of E at the surface is required to compute E. In practice, several steps and approximations are needed to obtain an explicit expression for $E(\vec{r})$ (see [142] p 503-504).

Vector scattering

A similar approach, based on the 3D Green's theorem, holds in the general case of 3D scattering, which is valid for electromagnetic scattering by the sea surface. We will obtain the famous *Stratton-Chu equations*, which form a coupled system for the total electric and magnetic fields (scattered + incident), denoted $\vec{E}(\vec{r}, t)$ and $\vec{B}(\vec{r}, t)$ respectively. They are again in baseband, but we allow a slow dependence in time as for the reflectivity in section 1.2.4. The Stratton-Chu equations [132] yield:

$$
\begin{aligned}
\vec{E}(\vec{r}, t) &= \vec{E}_{in}(\vec{r}) + \int_{S(t)} [i\omega\vec{n}(\vec{x}') \wedge \vec{B}(\vec{x}')G(\vec{r}, \vec{x}') + \vec{n}(\vec{x}') \wedge \vec{E}(\vec{x}') \wedge \nabla' G(\vec{r}, \vec{x}') \\
&+ \vec{n}(\vec{x}').\vec{E}(\vec{x}')\nabla' G(\vec{r}, \vec{x}')]\mathrm{d}S(\vec{x}') \quad (1.65) \\
\vec{B}(\vec{r}, t) &= \vec{B}_{in}(\vec{r}) - \int_{S(t)} [i\omega\mu_0\epsilon\epsilon_0\vec{n}(\vec{x}') \wedge \vec{E}(\vec{x}')G(\vec{r}, \vec{x}') - \vec{n}(\vec{x}') \wedge \vec{B}(\vec{x}') \wedge \nabla' G(\vec{r}, \vec{x}') \\
&- \vec{n}(\vec{x}').\vec{B}(\vec{x}')\nabla' G(\vec{r}, \vec{x}')]\mathrm{d}S(\vec{x}'). \quad (1.66)
\end{aligned}
$$

ϵ is the ratio of the squared wavenumber in the atmosphere (k_1) and sea (k_2): $\epsilon = k_2^2/k_1^2$. It is in fact assumed that $k_1 = \omega^2\epsilon_0\mu_0$ (the atmosphere is approximately vacuum). G is the 3D Green's function, which solves the 3D equivalent of equation (1.63). Solving the Stratton-Chu equations is not easy, and one often has to make approximations (see [113] or [41] for an advanced survey), or to use numerical methods [124]. If it was straightforward to solve the Stratton-Chu equation, we would still need to know $S(t)$, or equivalently the function $h(x, y, t)$ which describes the sea surface height as a function of time and horizontal coordinates. In practice, it is very challenging. The sea surface is an interface between the atmosphere and the sea and knowing it exactly would require to solve the complete ocean-atmosphere fluid dynamics. Using stereo imaging techniques, we can now measure (to some extent) the sea surface topography $S(t)$ simultaneously with its radar reflectivity (see [22]). It is a promising breakthrough for a better understanding of the sea surface reflectivity as a function of $S(t)$.

A simpler point of view is to describe it by a spectrum (Pierson-Moskowitz [112], Elfouhaily [40], see also [100]). It models well the idea that the sea surface is a superposition of waves of a range of wavelengths, from capillary waves (centimeters) to swell waves (hectometers) and beyond (see figure 1.13).

Figure 1.13: Scenic view of the sea surface in New Zealand.

It is legitimate to ask ourselves if we really need to know the exact time series of the sea surface reflectivity (or scattered field). In fact, we still have *a time series signal processing point of view*, and from this perspective, starting from the Stratton-Chu equations does not seem very relevant or efficient. We prefer to start from a much simpler model, the random walk model, which leads to statistical distributions of the sea clutter readily exploitable in signal processing of a time series.

1.3 The random walk model

We have introduced the complex reflectivity in section 1.2.4. In section 1.2.6, we saw that the complex reflectivity of the sea surface looked like noise. Even though it looks so, it does not mean that there is no structure in sea clutter. In this section, we present the random walk model. We will see that building on the discrete scatterers point of view, it leads to probability distributions for the sea clutter (or functions of it) at a fixed time t. For this whole section, refer to [70] chapters 1 to 4 and [141], [67], [71], [72] to a lesser extent.

1.3.1 Sketch of proof and the K distribution

From now on, we see the reflectivity as a random process, *i.e.* a time series of random variables. We denote it Ψ_t. In practice, each t corresponds to a pulse to which a reflectivity Ψ_t is attributed. This should lead to a discrete time series with timestep $\frac{1}{PRF}$ (Pulse Repetition Frequency) since the pulses cannot be infinitely close. In theory, we imagine that the reflectivity exists for all $t \geq 0$ (continuous time).

The random walk model starts from the fundamental idea that the reflectivity is a sum

of contributions over a population of independent scatterers, *i.e.* $\forall t \geq 0$:

$$\Psi_t = \sum_{n=1}^{N_t} a_t^{(n)} e^{i\phi_t^{(n)}}. \tag{1.67}$$

$a_t^{(n)}$ and $\phi_t^{(n)}$ are respectively the amplitude and phase of the n-th scatterer. It is assumed that for fixed t, the amplitudes $a_t^{(n)}$ are independent and identically distributed (i.i.d). Similarly, the phases $\phi_t^{(n)}$ are independent for different n, and all phases and amplitudes are independent. Finally, the number of scatterers N_t is itself a random variable. Moreover, all processes are stationary. Let $\bar{N} = \mathbb{E}[N_t]$ be the average number of scatterers. If we normalize the amplitudes by $\bar{N}^{1/2}$ and let $\bar{N} \to +\infty$, we obtain:

$$\Psi_t = \lim_{\bar{N} \to +\infty} \sum_{n=1}^{N_t} \frac{a_t^{(n)}}{\bar{N}^{1/2}} e^{i\phi_t^{(n)}}. \tag{1.68}$$

$\forall t$, Ψ_t is a complex random variable, whose properties shall be derived from those of $a_t^{(n)}$, N_t and $\phi_t^{(n)}$. We consider exclusively $\phi_t^{(n)}$ uniformily distributed over $[0, 2\pi[$. Also we denote:

$$\langle a^2 \rangle = \mathbb{E}[a_t^{(n)\,2}], \tag{1.69}$$

where $\mathbb{E}$ is the mathematical expectation (ensemble average). A good surprise is that knowing the whole distribution of $a_t^{(n)}$ is not necessary, $\langle a^2 \rangle$ is enough. Let $z_t = |\Psi_t|^2$ be the *intensity*. We consider now the two most common cases for N_t. The first case is $N_t = \mathbb{E}[N_t]$, *i.e.* there is a constant number of scatterers. In that case, it is shown in [70] that $\forall t$, z_t follows an exponential distribution:

$$p(z_t = z) = \frac{1}{\langle a^2 \rangle} e^{-\frac{z}{\langle a^2 \rangle}}. \tag{1.70}$$

This case is usually called *Gaussian* because the real and imaginary parts of the reflectivity $R_t = Re(\Psi_t)$ and $I_t = Im(\Psi_t)$ follow a joint Gaussian distribution:

$$p(R_t = R, I_t = I) = \frac{1}{\pi\langle a^2 \rangle} e^{-\frac{(R^2 + I^2)}{\langle a^2 \rangle}}. \tag{1.71}$$

The second usual case considers that $\forall t$, N_t is negative binomial distributed:

$$p(N_t = N) = \binom{N + \alpha - 1}{N} \frac{(\langle N \rangle / \alpha)^N}{(1 + \langle N \rangle / \alpha)^{N+\alpha}}, \tag{1.72}$$

with $\langle N \rangle = \mathbb{E}[N_t]$. In that case, we can show that z_t is K-distributed:

$$p(z_t = z) = \frac{2z^{\frac{\alpha-1}{2}}}{\Gamma(\alpha)} \left(\frac{\alpha}{\langle a \rangle^2} \right)^{\frac{\alpha+1}{2}} K_{\alpha-1}\left(2\sqrt{\frac{\alpha z}{\langle a^2 \rangle}} \right), \tag{1.73}$$

where K is a modified Bessel function of the second kind and $\langle a \rangle = \mathbb{E}[a_t^{(n)}]$. Also, when N_t follows a negative binomial distribution, the variance is written as:

$$var \frac{N_t}{\langle N \rangle} = \frac{1}{\langle N \rangle} + \frac{1}{\alpha}, \tag{1.74}$$

such that the asymptotic variance of $N_t/\langle N\rangle$ is $1/\alpha$ as $\langle N\rangle \to +\infty$. When $\alpha \to +\infty$, the variance goes to 0 and we get back to the Gaussian case. The K distribution is therefore usually prefered for the reflectivity of the sea surface. It is more general and empirically confirmed than the Gaussian model [141]. It is interesting to note that the random walk model can be used in lots of different contexts. We simply quote Jakeman and Ridley in [70] (see also [69] where several other applications are cited, such as stellar scintillation etc):

"*It (the random walk model) is an essentially exact model for light scattering by small particles. [...] the model is equally relevant to microwave scattering from raindrops or electron scattering from atomic defects, and it will be demonstrated in later chapters that it also provides a good representation for many aspects of scattering by continuous systems, such as rough surfaces and turbulent media.*"

An undeniable strengh of the random walk model is that it leads to probability distributions, like the K distribution for the intensity z_t, which can be used directly for example for target detection (see [141] chapter 6). We let the reader imagine how hard it would be to obtain similar statistics starting from the Stratton-Chu equations!

1.3.2 Limitations of the random walk model

We have seen that from the random walk model we can derive the distributions of functions of Ψ_t, like z_t, for one fixed time t. However, something crucial is missing in this formulation: the relation between Ψ_{t_1} and Ψ_{t_2}, for two subsequent times $t_1 < t_2$, is not specified. Only the static distributions (for a fixed t) are known. For example, how can we compute the moment:

$$\langle z_{t_1} z_{t_2} \rangle \tag{1.75}$$

in the framework of the random walk model? More generally, given $t_1 < t_2 \cdots < t_n$ and functions $f_1, f_2, \ldots, f_n$, how do we compute:

$$\langle f_1(\Psi_{t_1}) f_2(\Psi_{t_2}) ... f_n(\Psi_{t_n}) \rangle, \tag{1.76}$$

i.e. any moment with more than one time? Even more generally, how do we compute joint distributions with more than one time (from which moments can be computed), *e.g.* $p(z_{t_1} = z_1, z_{t_2} = z_2)$? In the same vein, we will see later on (for example chapter 4) that transition probabilities, for example:

$$p(z_{t_2} = z_2 \,|\, z_{t_1} = z_1), \tag{1.77}$$

are very important, and cannot be computed with the random walk model left alone. In general, the random walk model as in [70] and presented in the present section, cannot solve these problems related to *dynamics*. This is the first limitation. One can partially dodge the problem by specifying correlation times and designing *ad hoc* numerical procedures for simulating individual trajectories of the process (see [141] chapter 5). However, there is a much more satisfying and powerful framework: Field's model, presented in chapter 3.

A second limitation is of course its simplicity. We can question the axioms of the random walk model at the light of what the sea surface really is. Is the discrete scatterers model relevant? If so, are the scatterers independent? These questions will be even more pervasive under Field's model (see section 3.2.3).

1.3.3 Aim of the thesis

We have come to the point where the reader can understand the *raison d'être* of this thesis. Be it for oceanography, maritime surveillance or other purposes, there is a need for a good sea clutter model. Computations of the sea clutter through integrals over the sea surface (Stratton-Chu equations) are very limited to have 'statistical knowledge' of the sea clutter.

The random walk model, which reduces the sea surface to a family of point scatterers with random amplitudes and phases, is much more tractable and leads quickly to probability densities for various quantities such as the intensity. However, it is essentially *static* and lacks deepness to account for the *dynamics* of the sea clutter. Understanding these dynamics is not only inherently attractive: it has implications in terms of signal processing and applications for radar imagery or target detection (chapters 4 to 7). We propose to explore Field's model [48], which expresses the reflectivity of a random medium (*e.g.* the sea surface) as a solution to a *stochastic differential equation*. Field's model builds on the random walk model and precisely gives to the reflectivity its temporal structure, the dynamics which are absent in the random walk model.

The core of the thesis (chapters 3 to 7) is to present, clarify and build upon Field's model. If electromagnetic scattering modelling represents one layer of theory, developing signal processing tools is a second layer. We admit the random walk model and Field's model for the first layer, and construct (or keep constructing) the second one.

Radar remote sensing of the sea surface is one example of electromagnetic scattering from a random medium. The random walk model, and Field's model, are not specific to the sea surface (remember the glittering stars). This is one of their weakness (lack of specificity) but also one of their strength. Indeed, all the results presented in this work not only apply to sea clutter, but potentially to the complex reflectivity of any dynamic random medium. Of course, this holds within the limits of applicability of the model.

Chapter 2

Elements of stochastic analysis

Before introducing the model proposed by T. R. Field for the sea (or any random medium) clutter, which is a stochastic model, it is necessary to define some mathematical notions. We present in this chapter the concepts and tools that will be mainly used throughout this thesis. In section 2.1, we present different types of stochastic processes and define several modes of convergence for random variables and random processes. In section 2.2, we define Itô's integral of a stochastic process and stochastic differential equations. We introduce basic numerical schemes to solve them. In section 2.3, we focus on the concept of transition probabilities since it is a recurrent topic in chapters 3, 4 and 5. In section 2.4, we define the covariation of processes and basic results about volatility estimation from a time series, which will be usefull in chapter 7. We briefly compare the maximum likelihood and volatility-based estimators in section 2.5. Section 2.6 presents the Wiener-Khinchin theorem, and section 2.7 concludes.

We think that the reader could either skip (or skim) parts of this chapter and come back to them when he needs to, or read the whole chapter linearly and keep in mind that every notions and results which are presented are useful to the understanding of the remaining of this thesis. In particular, section 2.1 includes a lot of definitions. A reader familiar with the brownian motion, martingales, Markov processes etc can skip this part. Section 2.2 is quite technical with the objective to present rigorously stochastic differential equations. The reader uninterested in mathematical considerations can jump directly to section 2.2.3 which, by the way of Euler-Maruyama's scheme, gives meaning and intuition on stochastic differential equations. The remaining sections are more short and digestible.

2.1   Stochastic processes

In this section, we need to define some fundemental concepts without which we will not be able to work with stochastic differential equations. We define mostly brownian motions, (semi)martingales, Markov processes and stochastic convergence. These are fundamental because stochastic differential equation requires Itô integration with respect to a brownian motion. Itô integration requires a limiting procedure, for which one must have defined convergence. We will also see that a solution to a stochastic differential equations is a Markov process, and a semimartingale. These are two important results, some consequences of which are given in section 2.3 and 2.4.

2.1.1 Gaussian processes

As a preliminary step, we remind a few definitions that can be found in any textbook about probability theory.

Definition 2.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space (*i.e.* a measurable space with a positive measure $\mathbb{P}$ such that $\mathbb{P}(\Omega) = 1$) and $(E, \mathcal{E})$ a measurable space. A *random variable* is an $\mathcal{F}$-measurable application from Ω to E:

$$
\begin{aligned}
X: \quad \Omega &\to E \\
\omega &\mapsto X(\omega)
\end{aligned}
$$

The application is said to be $\mathcal{F}$-*mesurable* if $\forall\, B \in \mathcal{E},\; X^{-1}(B) \in \mathcal{F}$.

In this first definition, the space E and its σ-algebra $\mathcal{E}$ are not specified. When $E = \mathbb{R}$ with its Borel σ-algebra $B(\mathbb{R})$, we say that X is a *real-valued random variable* or *one-dimensional random variable* or simply *random variable* where it is implicit that it is real-valued. When $E = \mathbb{R}^n$ with its Borel σ-algebra $B(\mathbb{R}^n)$, we say that X is a random vector. We remind that the Borel σ-algebra of a topological space is the σ-algebra generated by its topology.

Definition 2.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X_0 a real-valued random variable. X_0 is a *Gaussian random variable* of mean μ and variance σ^2 if its law has the density probability (or distribution):

$$
\begin{aligned}
f_{\mu,\sigma}: \quad \mathbb{R} &\to \mathbb{R} \\
x &\mapsto f(x) = \frac{1}{\sigma\sqrt{2\pi}}\; e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}
\end{aligned}
$$

We remind that the law of a random variable X is the direct image by X of the probability measure $\mathbb{P}$. It is usually denoted $\mathbb{P}_X$. It is a probability measure on the measurable space $(E, \mathcal{E})$, and by definition $\forall\, B \in \mathcal{E}, \mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B)$. Random variables or random vectors which have a density probability relative to Lebesgue's measure are said to be absolutely continuous. In this thesis, we stay exclusively in the absolutely continuous framework. In that case, if X is an $\mathbb{R}^n$-valued random vector with probability density $x \mapsto f(x)$, $x \in \mathbb{R}^n$, and g a measurable application from $\mathbb{R}^n$ to $\mathbb{R}^m$, we can compute any probability of $g(X)$ by direct integration:

$$
\mathbb{P}(g(X) \in B) = \mathbb{P}(X \in g^{-1}(B)) = \int_{g^{-1}(B)} f(x) d\lambda_n(x), \tag{2.1}
$$

where $B \in B(\mathbb{R}^m)$, and λ_n is Lebesgue measure on $\mathbb{R}^n$.

We can extend the concept of a Gaussian random variable to that of a Gaussian random vector:

Definition 2.3. (Gaussian random vector) Let X be a random vector,

$$
\begin{aligned}
X: \quad \Omega &\to \mathbb{R}^n \\
\omega &\mapsto X(\omega)
\end{aligned}
$$

with $X(\omega) = \begin{bmatrix} X_1(\omega) & X_2(\omega) & \ldots & X_n(\omega) \end{bmatrix}^\top$.

X is a *Gaussian random vector* if $\forall\, a = \begin{bmatrix} a_1 & a_2 & \ldots & a_n \end{bmatrix}^\top \in \mathbb{R}^n$, the random variable $\sum_{i=1}^{n} a_i X_i$ is a Gaussian random variable.

The expectation of X is $m_X = \begin{bmatrix} \mathbb{E}[X_1] & \mathbb{E}[X_2] & \ldots & \mathbb{E}[X_n] \end{bmatrix}^\top$ and its covariance matrix Σ_X is:

$$\Sigma_X = \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \ldots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \cdots & \cdots & \mathrm{Var}(X_n) \end{bmatrix}. \tag{2.2}$$

It holds that for a Gaussian vector X, m_X and Σ_X determine entirely the law of X. If Σ_X is invertible (*i.e.* non singular), X is absolutely continuous and we have:

$$p_X(x_1, x_2, ..., x_n) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_X|}} \exp\left(-\frac{1}{2}(\mathbf{x} - m_X)^\top \Sigma_X^{-1} (\mathbf{x} - m_X)\right), \tag{2.3}$$

where $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix}^\top$ and $|\Sigma_X| = \det(\Sigma_X)$.

We can generalize the concept of Gaussian vector to define that of a Gaussian (stochastic) process. A *stochastic process* $X = (X_t)_{t \in T}$ with T a subset of $\mathbb{R}^+$, is a collection of random variable (or vector), *i.e.* $\forall t \in T$, X_t is a random variable from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathbb{R}^n, B(\mathbb{R}^n))$. The notation T is chosen for the interval because most of the time T is simply the interval $[0, T]$ where $T \in \mathbb{R}^+$, or $[0, T[$ if $T = +\infty = \mathbb{R}^+$. When the interval is not specified, we may use the notation $(X_t)_t$, or even simply X_t (especially from chapter 4 on).

Definition 2.4. (Gaussian process) A process $X = (X_t)_{t \in T}$ where X_t takes values in $\mathbb{R}^n$ is *Gaussian* if $\forall k \in \mathbb{N}$ and $\forall t_1, t_2, ..., t_k \in T$, the vector $\begin{bmatrix} X_{t_1} & X_{t_2} \ldots & X_{t_k} \end{bmatrix}^\top$ is a Gaussian vector. It is a centered Gaussian process if also $\forall t \in T$, $\mathbb{E}[X_t] = \mathbf{0}$.

2.1.2 The Brownian motion

We now turn to one of the basic components of stochastic differential equations, and therefore of the model for the sea clutter we use throughout this thesis: the brownian motion. There are several equivalent definitions to define the brownian motion. We first define a $\mathbb{R}$-valued brownian motion. Let K be the covariance function of a process X. We assume that $T = \mathbb{R}^+$. By definition:

$$\begin{aligned} K: \quad (\mathbb{R}^+)^2 &\to \mathbb{R} \\ (s, t) &\mapsto K(s, t) \equiv Cov(X_s, X_t) \end{aligned}$$

Definition 2.5. (Brownian Motion) Let $(B_t)_{t \geq 0}$ be a real-valued stochastic process. $(B_t)_{t \geq 0}$ is a *brownian motion* if and only if:

   i. it is a centered Gaussian process,

   ii. its trajectories are continuous, *i.e.* $\forall \omega \in \Omega$, $t \mapsto B_t(\omega)$ is continuous,

   iii. its covariance function is $K(s, t) = min(s, t)$.

In fact, as long as the process is Gaussian and the covariance is $K(s,t) = min(s,t)$, Kolmogorov's Lemma enables us to assume that the trajectories are continuous. More precisely, if $(B_t)_t$ is such a process, $\exists\,(\tilde{B}_t)_t$ such that $\mathbb{P}(\forall\, t \geq 0, B_t = \tilde{B}_t) = 1$ and such that $(B_t)_t$ has locally holderian trajectories with exponent $\frac{1}{2} - \delta$ for all $\delta \in]0, \frac{1}{2}[$ (see [54] p 24).

There are at least 2 more definitions for the brownian motion, which are equivalent to definition 2.5 as stated by the next proposition.

Proposition 2.1. *Let $(B_t)_{t\geq 0}$ be a continuous stochastic process. It then holds that the 3 following properties are equivalent:*

   i. $(B_t)_{t\geq 0}$ is a brownian motion,

  ii. $X_0 = 0$ a.s. and $\forall\, 0 \leq s < t$, $X_t - X_s$ is independent of $\sigma(X_r, r \leq s)$ (σ-algebra generated by the random variables X_s for $s \leq t$) and follows the law $\mathcal{N}(0, t-s)$,

 iii. $X_0 = 0$ a.s. and $\forall\, 0 = t_0 \leq t_1 \leq ... \leq t_p$, the increments $X_{t_i} - X_{t_{i-1}}$ with $i = 1, 2, \ldots, p$ are independent and $X_{t_i} - X_{t_{i-1}}$ follows the law $\mathcal{N}(0, t_i - t_{i-1})$.

The abbreviation *a.s.* means *almost surely*, or *with probability one*. Therefore, a brownian motion is a continuous stochastic process (*i.e.* continuous trajectories) which starts a.s. from 0, with increments $X_t - X_s$ independent from the past and which are Gaussian distributed with mean 0 and variance $t - s$. It is the essential properties to get an intuition of what a brownian motion is.

Definition 2.6. (Multidimensional Brownian Motion) A *multidimensional brownian motion* or *k-dimensional brownian motion*, is a vector-valued stochastic process:

$$B_t = \begin{bmatrix} B_t^{(1)} & B_t^{(2)} & \ldots & B_t^{(k)} \end{bmatrix}^\top \tag{2.4}$$

such that for all $i = 1, 2, \ldots, k$, $(B_t^{(i)})_t$ is a brownian motion, and such that for all $i \neq j$, $(B_t^{(i)})_t$ and $(B_t^{(j)})_t$ are independent stochastic processes.

The independence property required in definition 2.6 means that for all $i \neq j$, for all random vector V_i, V_j extracted from $(B_t^{(i)})_t$, $(B_t^{(j)})_t$ respectively, V_i and V_j are independent. The brownian motion has many properties that we cannot enumerate. Refer to [27] and [101] for precise statements. We shall cite at least one key properties of the brownian motion: its trajectories are continuous, but a.s., its trajectories are nowhere differentiable. The brownian motion is therefore irregular at all scales. A.s., a brownian motion trajectory has Haussdorff dimension 3/2.

Let $C(\mathbb{R}^+, \mathbb{R})$ be the space of continuous functions from $\mathbb{R}^+$ to $\mathbb{R}$. A $\mathbb{R}$-valued process X can be seen as an application:

$$\begin{aligned} X : \quad &\Omega \to C(\mathbb{R}^+, \mathbb{R}) \\ &\omega \mapsto X(\omega) \end{aligned} \tag{2.5}$$

where $X(\omega)$ is the whole trajectory $t \mapsto X_t(\omega) = X(t, \omega)$. One can endow $C(\mathbb{R}^+, \mathbb{R})$ with a topology defined by the uniform convergence over all compact sets. It is then possible to define the Borel σ-algebra on $C(\mathbb{R}^+, \mathbb{R})$ which becomes a measurable space. We define the *Wiener measure* W on $C(\mathbb{R}^+, \mathbb{R})$ as the image measure of $\mathbb{P}$ by the application B where B is a brownian motion. One can show that W does not depend on the brownian motion, *i.e.* if B' is another brownian motion, then for all measurable set A of the space $C(\mathbb{R}^+, \mathbb{R})$ we have $\mathbb{P}((B_t)_t \in A) = \mathbb{P}((B_t)'_t \in A)$. The law of a brownian motion is therefore unique and we can define a canonical brownian motion (see [54] p 25). That is why we alternatively speak of *a* brownian motion or of *the* brownian motion.

2.1.3 Markov processes

There are many different ways to define a Markov process $(X_t)_t$ depending on the required level of abstraction. There are at least 4 possibilities: discrete time or continuous time, discrete or continuous state space, the state space being the measurable space $(E, \mathcal{E})$ in which X_t takes its values for all t. We consider here only continous time and continuous state space ($\mathbb{R}$ or $\mathbb{R}^n$).

Let X and Y be random variables (or vectors) defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and A a Borel set. We define:

$$\mathbb{P}(X \in A | Y) = \mathbb{E}[I_A(X) | Y], \tag{2.6}$$

where I_A is the *indicator function* of the set A, and $\mathbb{E}[I_A(X) | Y]$ is the *conditional expectation* of $I_A(X)$ with respect to Y. We think that it is useful to precise this by giving a definition. This definition is not comprehensive to have a good understanding of conditional expectation, but should be enough for the use we make of it (see [85] for a detailed exposition).

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $L^2(\Omega, \mathcal{F}, \mathbb{P})$ the space of squared-integrable measurable applications from Ω to $\mathbb{R}$. It is a Hilbert space, and if $U, V \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, the scalar product reads:

$$< U, V >= \mathbb{E}[UV] \tag{2.7}$$

and defines a corresponding norm, and thus distance on L^2 which is thusly a metric space.

Definition 2.7. (Conditional expectation) Let $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ and B a sub-σ-algebra of $\mathcal{F}$. We name *conditional expectation of X relative to B* the orthogonal projection of X on the sub-space $L^2(\Omega, B, \mathbb{P})$. We denote it $\mathbb{E}[X | B]$.

For any two random variables U and V, the conditional expectation of U with respect to V, $\mathbb{E}[U | V]$ is a random variable, not just a number or vector like the usual expectation of a random variable. $\mathbb{E}[U | V]$ can be thought of as the projection of the random variable U on a subspace of random variables which 'takes into account' the random variable V. In this subspace, it is the closest random variable to U.

If Y is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ we set:

$$\mathbb{E}[X | Y] = \mathbb{E}[X | \sigma(Y)], \tag{2.8}$$

where we recall that $\sigma(Y)$ is the σ-algebra generated by Y (next definition). If we have several random variables $Y_1, Y_2, \ldots, Y_n$, then:

$$\mathbb{E}[X | Y_1, ..., Y_n] = \mathbb{E}[X | \sigma(Y_1, ..., Y_n)]. \tag{2.9}$$

Definition 2.8. (Generated σ-algebra) Let $(\Omega, \mathcal{F})$ be a measurable space and $Y_1, Y_2, \ldots, Y_n$ be n measurable applications. The σ-algebra generated by $Y_1, Y_2, \ldots, Y_n$ is the smallest σ-algebra on Ω such that all the Y_i are measurable relative to it.

Definition 2.9. (Markov process) Let $X = (X_t)_{t \in T}$ be a stochastic process with values in $(E, \mathcal{E})$. We say that X is a *Markov process* if $\forall s < t$ and $\forall A \in \mathcal{E}$, it holds:

$$\mathbb{P}(X_t \in A | X_u, u \leq s) = \mathbb{P}(X_t \in A | X_s). \tag{2.10}$$

This definition, from [55] chapter 3, enables us to define a bunch of concepts like transition kernels, semi-groups, Markov-Feller processes, infinitesimal generator etc. We do not need such a formalism. We turn to a more simple definition of a Markov process (see [56] for example). First, we assume that we are in the framework of absolute continuity, *i.e.* all extracted random vectors $X_{t_1}, X_{t_2}, \ldots, X_{t_k}$ from the stochastic process $(X_t)_t$ are absolutely continuous with respect to the Lebesgue measure of the space they live in ($\mathbb{R}^{kn}$ if for all t X_t lives in $\mathbb{R}^n$).

Definition 2.10. (Joint distribution) Let $\left[X_1, X_2, \ldots, X_n\right]^\top$ be a $\mathbb{R}^n$-valued random vector. If it is absolutely continuous with respect to Lesbegue measure λ_n, its joint distribution is a function:

$$
\begin{aligned}
p: \quad & \mathbb{R}^n \to \mathbb{R}^+ \\
& (x_1, \ldots, x_n) \mapsto p(X_1 = x_1, \ldots, X_n = x_n)
\end{aligned}
\tag{2.11}
$$

such that for all measurable set A in $B(\mathbb{R}^n)$:

$$
\mathbb{P}(\left[X_1, X_2, \ldots, X_n\right]' \in A) = \int_A p(X_1 = x_1, \ldots, X_n = x_n) d\lambda_n(x_1, \ldots, x_n).
\tag{2.12}
$$

Definition 2.11. (Conditional probabilities) Let $Y = (Y_t)_{t \in T}$ be a random process. Let us choose k times $t_1 < t_2 < \cdots < t_k$ for which we impose the condition $Y_{t_1} = y_1, \ldots, Y_{t_k} = y_k$, where $y_i \in \mathbb{R}^n$ for all i. For m times $t_{k+1} < t_{k+2} \cdots < t_{k+m}$, we define the joint distribution of $\left[Y_{t_{k+1}}, Y_{t_{k+2}}, \ldots, Y_{t_{k+m}}\right]^\top$ given the condition $Y_{t_1} = y_1, \ldots, Y_{t_k} = y_k$ by

$$
p(Y_{t_{k+1}} = y_{k+1}, \ldots, Y_{t_{k+m}} = y_{k+m} \mid Y_{t_1} = y_1, \ldots, Y_{t_k} = y_k) = \frac{p(Y_{t_1} = y_1, \ldots, Y_{t_{k+m}} = y_{k+m})}{p(Y_{t_1} = y_1, \ldots, Y_{t_k} = y_k)}.
\tag{2.13}
$$

This last definition is of course the Bayes formula. We can now give a much simpler and intuitive definition of a Markov process for continuous time and taking its values in $\mathbb{R}^n$.

Definition 2.12. (Markov process) Let $Y = (Y_t)_{t \in T}$ be an absolutely continous stochastic process. We say that Y is a *Markov process* if $\forall k \in \mathbb{N}$, $\forall t_1 < t_2 < \ldots < t_k$, we have:

$$
p(Y_{t_k} = y_k \mid Y_{t_1} = y_1, \ldots, Y_{t_{k-1}} = y_{k-1}) = p(Y_{t_k} = y_k \mid Y_{t_{k-1}} = y_{k-1}).
\tag{2.14}
$$

Let t_0 be the time origin of the process (often taken at 0). If it is also true that $\forall s < t$:

$$
p(Y_t = y \mid Y_s = x) = p(Y_{t_0 + t - s} = y \mid Y_{t_0} = x),
\tag{2.15}
$$

then we say that Y is a *homogeneous Markov process*.

Definition 2.12 means that a Markov process depends only on the most recent past. The memory of the past before the most recent past is unimportant. From now on, we shall think of notations like $p(X = x)$ as the 'probability that the random variable X takes the value x', though since is not strictly accurate.

2.1.4 Martingales

In this section, we will define the notion of martingale. The main interest of introducing martingales in this thesis comes from its use to define Itô integrals of stochastic process. In particular, solutions to stochastic differential equations are semimartingales in the general case, and martingales if the drift vanishes. These properties will be useful in section 2.4. We have to define a bunch a words that are necessary for a proper definition of Itô integral.

Definition 2.13. (Filtration) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *filtration* is an increasing family $(\mathcal{F}_t)_{0 \leq t \leq +\infty}$ of sub-σ-algebras of $\mathcal{F}$.

If $s \leq t$, all $\mathcal{F}_s$-measurable set is also $\mathcal{F}_t$-measurable, and a $\mathcal{F}_s$-measurable application $X : \Omega \mapsto E$ is also $\mathcal{F}_t$-measurable.

Definition 2.14. (Canonical filtration of a process) Let $(X_t)_{t \geq 0}$ be a stochastic process defined on the probability space$(\Omega, \mathcal{F}, \mathbb{P})$. We call *canonical filtration* (understood 'relative to $(X_t)_{t \geq 0}$') the filtration on $(\Omega, \mathcal{F}, \mathbb{P})$ defined by:

$$\mathcal{F}_t = \sigma(X_s, s \leq t). \tag{2.16}$$

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ endowed with a filtration $(\mathcal{F}_t)_{0 \leq t \leq +\infty}$ is a *filtered* probability space.

Definition 2.15. (Progressive σ-algebra) Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq +\infty}, \mathbb{P})$ be a filtered probability space. The *progressive σ-algebra* is the class $\mathcal{P}$ of set of $\Omega \times \mathbb{R}^+$ defined by:

$$\mathcal{P} = \{A \subseteq \Omega \times \mathbb{R}^+ | \forall t \geq 0, A \cap (\Omega \times [0,t]) \in \mathcal{F}_t \times B[0,t]\}. \tag{2.17}$$

We now define a few adjectives that we can use to qualify a stochastic process and to define the important notion of semimartingale.

Definition 2.16. (Measurable process) Let $X = (X_t)_{t \geq 0}$ be a stochastic process defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. X can be seen as an application:

$$
\begin{aligned}
X : \quad \Omega \times \mathbb{R}^+ &\to E \\
(\omega, t) &\mapsto X_t(\omega)
\end{aligned}
$$

where E is a measurable space ($\mathbb{R}$ or $\mathbb{R}^n$ for us). We say that X is a *measurable process* if it is a measurable application relative to the σ-algebra on E and to the product σ-algebra $\mathcal{F} \times B(\mathbb{R}^+)$ on $\Omega \times \mathbb{R}^+$.

Definition 2.17. (Adapted process) Let $X = (X_t)_{t \geq 0}$ be a stochastic process defined on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq +\infty}, \mathbb{P})$. We say that X is *adapted* (understood 'relative to $(\mathcal{F}_t)_{0 \leq t \leq +\infty}$') if $\forall t \geq 0$, the random variable X_t is $\mathcal{F}_t$-measurable.

Definition 2.18. (Progressive process) Let $X = (X_t)_{t \geq 0}$ be a stochastic process defined on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq +\infty}, \mathbb{P})$, and $\mathcal{P}$ the progressive σ-algebra on $\Omega \times \mathbb{R}^+$. We say that X is *progressive* or *progressively measurable* if it is measurable relative to $\mathcal{P}$. It is equivalent to saying that for all t, the mapping:

$$
\begin{aligned}
X : \quad \Omega \times [0,t] &\to E \\
(\omega, s) &\mapsto X_s(\omega)
\end{aligned}
$$

is measurable relative to $\mathcal{F}_t \times B([0,t])$ where $B([0,t])$ is the Borel σ-algebra on $[0,t]$.

We now assume that our stochastic processes takes values in $\mathbb{R}$.

Definition 2.19. (Martingale) Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \le t \le +\infty}, \mathbb{P})$ be a filtered probability space. Let $X = (X_t)_{t \ge 0}$ be a stochastic process, adapted to $(\mathcal{F}_t)_{0 \le t \le +\infty}$. We assume that $\forall\, t \ge 0$, $X_t \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Then X is a *martingale* if:

$$\forall\, 0 \le s < t,\ \mathbb{E}[X_t|\mathcal{F}_s] = X_s. \tag{2.18}$$

Definition 2.20. (Finite variation process) Let $f : \mathbb{R}^+ \to \mathbb{R}$ be a function. The variation of f on $[0,t]$ is $sup_{\{\Delta_t\}}\{\sum_{k=0}^{n-1} |f(t_{k+1}) - f(t_k)|\}$ taken on all subdivisions Δ_t of $[0,t]$. f has *finite variation* if $\forall\, t \ge 0$, $sup_{\{\Delta_t\}}\{\sum_{k=0}^{n-1} |f(t_{k+1}) - f(t_k)|\} < +\infty$. Let $(A_t)_t$ be a stochastic process. We say that A is a *finite variation* process if a.s., the trajectory $t \mapsto A_t$ has finite variation.

Definition 2.21. (continuous semimartingale) Let $(X_t)_t$ be a continuous and adapted process. It is a *semimartingale* if there exists a decomposition of X:

$$\forall\, t \ge 0,\ X_t = A_t + M_t, \tag{2.19}$$

where $(M_t)_t$ is a continuous local martingale and $(A_t)_t$ an adapted, continuous, finite variation on compacts of $\mathbb{R}^+$ process, with $A_0 = 0$.

Semimartingales are very important in the theory of stochastic differential equations, because their solutions are precisely semimartingales (see proposition 2.2). The finite variation process correspond to a process which variations are limited: it will come from the deterministic component of the stochastic differential equations (drift), while the volatility part of the stochastic differential equations, which is the true innovation compared to ordinary differential equations, has infinite variation over any interval. This is due to the presence of the brownian motion and its infinite variation. We will not define local martingales since it requires another notion, stopping times, which is not necessary in this work. We just add that all martingales are in particular local martingales.

2.1.5   Convergence and stochastic Landau notations

We first define the most classical type of convergence for sequences of random variables. Convergence is needed to define Itô integral in the next section, as well as limit theorems for volatility-based estimators in section 2.4. Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of random vectors and X a random vector taking values in $(\mathbb{R}^n, B(\mathbb{R}^n))$, all defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 2.22. (Convergence in distribution) We say that $(X_n)_{n \in \mathbb{N}}$ *converges in distribution* to X if $\forall\, A \in B(\mathbb{R}^n)$,

$$\lim_{n \to +\infty} \mathbb{P}(X_n \in A) = \mathbb{P}(X \in A). \tag{2.20}$$

A must be a continuity set of X, *i.e.* $\mathbb{P}(X \in \partial A) = 0$ where ∂A is the boundary of A. For $n = 1$, the definition is given in terms of the cumulative distribution functions. Convergence in distribution is denoted $X_n \xrightarrow{d} X$. It is also called *weak convergence* because it is the weakest type of convergence we will define.

Definition 2.23. (Convergence in probability) We say that $(X_n)_{n \in \mathbb{N}}$ *converges in probability* to X if $\forall \, \epsilon > 0$:

$$\lim_{n \to +\infty} \mathbb{P}(\|X_n - X\| \geq \epsilon) = 0, \tag{2.21}$$

where $\|.\|$ is any norm on $\mathbb{R}^n$. Convergence in probability is denoted $X_n \xrightarrow{\mathbb{P}} X$.

Definition 2.24. (Almost sure convergence) We say that $(X_n)_{n \in \mathbb{N}}$ *converges almost surely* to X if

$$\mathbb{P}(\lim_{n \to +\infty} \|X_n(\omega) - X(\omega)\| = 0) = 1. \tag{2.22}$$

Almost sure convergence is denoted $X_n \xrightarrow{\text{a.s.}} X$.

Definition 2.25. (Convergence in mean) We say that $(X_n)_{n \in \mathbb{N}}$ *converges in the p-th mean* to X, where $p \geq 1$, if $\forall \, n \in \mathbb{N}$, $\mathbb{E}[\|X_n\|^p]$ exists, if $\mathbb{E}[\|X\|^p]$ exists, and if

$$\lim_{n \to \infty} \mathbb{E}[\|X_n - X\|^p] = 0. \tag{2.23}$$

Convergence in mean is denoted $X_n \xrightarrow{L^p} X$, because it is a convergence in the normed space L^p of random vectors Y such that:

$$\mathbb{E}[\|Y\|^p]^{1/p} = \left(\int_{\Omega} \|Y(\omega)\|^p \, d\mathbb{P}(\omega) \right)^{1/p} < \infty. \tag{2.24}$$

The left-hand side of equation (2.24) defines the L^p norm of Y.

We defined 4 types of convergence, and there is a well-known hierarchy between them: almost sure implies probability which implies distribution. In parallel, p-th mean implies probability which implies distribution. We understand why convergence in distribution is called weak convergence. There exists at least another type of convergence: stable convergence (see [12], [64]), which is in between convergence in probability and convergence in distribution. The definition is a bit tricky so we will skip it but it is useful for convergence theorems in the context of volatility estimation (section 2.4 and appendix A).

Finally, we define stochastic Landau notations, which will sometimes turn out to be useful.

Definition 2.26. (Little o) Let $(X_n)_{n \in \mathbb{N}}$ and $(\delta_n)_{n \in \mathbb{N}}$ be two sequences of random variables. We write $X_n = o_{\mathbb{P}}(\delta_n)$ if:

$$\frac{X_n}{\delta_n} \xrightarrow{\mathbb{P}} 0. \tag{2.25}$$

Definition 2.27. (Big O) Let $(X_n)_{n \in \mathbb{N}}$ and $(\delta_n)_{n \in \mathbb{N}}$ be two sequences of random variables. We write $X_n = O_{\mathbb{P}}(\delta_n)$ if $\forall \, \epsilon > 0$, $\exists \, M > 0$ such that:

$$\mathbb{P}\left(\frac{X_n}{\delta_n} > M \right) < \epsilon. \tag{2.26}$$

2.2 Stochastic Differential Equations

We are now in a position to define Itô integrals of stochastic processes and stochastic differential equations. To arrive at a proper definiton of stochastic differential equations, one has no choice but to dive more into abstract definitions. We do not want to dodge this difficulty and give here a detailed theoretical derivation. However, once stochastic differential equations are well defined, one can rely on Euler-Maruyama numerical scheme only (section 2.2.3) to both solve them and get a very good intuition of what they mean. Some portions of this section may therefore be seen as only temporarily relevant and may then be ignored in the rest of the thesis.

2.2.1 Itô integral

We present here a construction of Itô's integral limited to the case where we integrate relative to a brownian motion. We wish to keep the presentation short and clear. A more general presentation would show that one can integrate processes not only relative to brownian motions, but any process in a subclass of martingales (see [86]). From now on, our presentation mimics that of Oksendal in [105]. Let $(B_t)_{t\geq 0}$ be a brownian motion and $\mathcal{F}_t = \sigma(B_s, s \leq t)$ its canonical filtration. There is a link between σ-algebra and *information*, explicited by the following proposition (see [105] chapter 3).

Proposition 2.2. *Let $h : \Omega \to (\mathbb{R}, B(\mathbb{R}))$. h is $\mathcal{F}_t$-measurable if and only if it is the almost sure limit of a sum of functions of the form:*

$$g_1(B_{t_1})g_2(B_{t_2})...g_k(B_{t_k}), \tag{2.27}$$

where the functions g_i are continuous and bounded, and where $\forall\, i, t_i \leq t$.

Therefore, to evaluate a $\mathcal{F}_t$-measurable application, one only need the values of the brownian motion $(B_t)_{t\geq 0}$ before time t! This will help us understand $\mathcal{F}_t$-measurability conditions.

Definition 2.28. (Space $\mathcal{V}$) We define $\mathcal{V}$ as a class of processes defined by $f \in \mathcal{V} \Leftrightarrow f$ verifies the 3 following conditions:

i. $(t, \omega) \mapsto f(t, \omega)$ is a measurable process,

ii. $\omega \mapsto f(t, \omega)$ is $\mathcal{F}_t$-measurable,

iii. $\mathbb{E}\left(\int_0^T f(s, \omega)^2 \mathrm{d}s\right) < +\infty$.

We can then define the stochastic integral of $f \in \mathcal{V}$. To do so, we approximate f by an *elementary* processes in $\mathcal{V}$. An elementary process can be written:

$$H_t(\omega) = H(t, \omega) = \sum_{k=0}^{n-1} Z_k(\omega) I_{]t_k, t_{k+1}]}(t). \tag{2.28}$$

If it is in the class $\mathcal{V}$, then we can show that $\forall k$, $Z_k(\omega)$ is $\mathcal{F}_{t_k}$-measurable. We define the stochastic integral of such a process H in the following way:

Definition 2.29. (Itô integral of an elementary process) Let $H \in \mathcal{V}$ be an elementary process as in equation (2.28). The *Itô integral* of H between 0 and t is defined as:

$$\int_0^t H_s dB_s = \sum_{k=0}^{n-1} Z_k(B_{t_{k+1}, t} - B_{t_k, t}), \tag{2.29}$$

where $B_{t_k, t}$ must be understood as $B_{min(t_k, t)}$.

We now define the Itô integral of any stochastic process in $\mathcal{V}$ thanks to the 2 following propositions.

Proposition 2.3. *(Itô isometry) Let $H(t, \omega)$ be an elementary process in $\mathcal{V}$. Then it holds:*

$$\mathbb{E}\left[\left(\int_0^T H(t, \omega) \mathrm{d}B_t\right)^2\right] = \mathbb{E}\left[\int_0^T H(t, \omega)^2 \mathrm{d}t\right]. \tag{2.30}$$

This can be written more simply:

$$\left\|\int_0^T H(s, \omega) \mathrm{d}B_s\right\|_{L^2(\Omega)}^2 = \|H\|_{L_T^2(B)}^2. \tag{2.31}$$

$\int_0^T H(s, \omega) \mathrm{d}B_s$ is a random variable, so the left-hand side of equation (2.31) is just the L^2-norm of this random variable. However, H is a process and the right-hand side is the norm of a process which we define as:

$$\|H\|_{L_T^2(B)}^2 = \mathbb{E}\left[\int_0^T H(t, \omega)^2 \mathrm{d}t\right]. \tag{2.32}$$

The next proposition gathers three propositions in one.

Proposition 2.4. *The three following statements are true:*

i. *Let $f \in \mathcal{V}$. It holds that $\exists (h_n)_{n \in \mathbb{N}}$ a sequence of bounded processes of $\mathcal{V}$ such that $\mathbb{E}[\int_0^T (f - h_n)^2 \mathrm{d}s)] = \|f - h_n\|_{L^2(B)}^2 \xrightarrow{n \to +\infty} 0$.*

ii. *Let $f \in \mathcal{V}$ be a bounded process. It holds that $\exists (h_n)_{n \in \mathbb{N}}$ a sequence of bounded and continuous processes of $\mathcal{V}$ such that $\mathbb{E}[\int_0^T (f - h_n)^2 \mathrm{d}s)] = \|f - h_n\|_{L^2(B)}^2 \xrightarrow{n \to +\infty} 0$.*

iii. *Let $f \in \mathcal{V}$ be a bounded and continuous process. It holds that $\exists (h_n)_{n \in \mathbb{N}}$ a sequence of elementary processes of $\mathcal{V}$ such that $\mathbb{E}[\int_0^T (f - h_n)^2 \mathrm{d}s)] = \|f - h_n\|_{L^2(B)}^2 \xrightarrow{n \to +\infty} 0$.*

This proposition asserts that for every process f of $\mathcal{V}$, there is a sequence $(h_n)_{n \in \mathbb{N}}$ of elementary processes of $\mathcal{V}$ which converges to f in the sense of the norm $\|.\|_{L_T^2(B)}$. One can show that Itô isometry ensures that $\int_0^T h_n \mathrm{d}B_t$ is a Cauchy sequence and converges in the $L^2(\Omega)$ norm sense to a random variable.

Definition 2.30. (Itô integral of a process in $\mathcal{V}$) Let $f \in \mathcal{V}$. Let $(h_n)_{n \in \mathbb{N}}$ be a sequence of elementary processes of $\mathcal{V}$ which converges to f in the sense of the norm $\|.\|_{L_T^2(B)}$. Then the *Itô integral* of f over $[0, T]$ is defined as:

$$\int_0^T f \mathrm{d}B_t = \lim_{n \to +\infty} \int_0^T h_n \mathrm{d}B_t, \tag{2.33}$$

where the limit is taken in the $L^2(\Omega)$ norm sense.

We can actually show that $\int_0^T h_n \mathrm{d}B_t$ is a uniform in t Cauchy sequence for $t \in [0, T]$, and that as a result, $\int_0^T h_n \mathrm{d}B_t$ converges in norm $L^2(\Omega)$ uniformily in t for $t \in [0, T]$.

We can generalize a bit the definition 2.30 by changing the third condition in the definition 2.28 of $\mathcal{V}$. To do so, we define a new class of processes: $\mathcal{W}$.

Definition 2.31. (Space $\mathcal{W}$) We define $\mathcal{W}$ as a class of processes defined by $f \in \mathcal{W} \Leftrightarrow f$ verifies the 3 following conditions:

i. $(t, \omega) \mapsto f(t, \omega)$ is a measurable process,

ii. $\omega \mapsto f(t, \omega)$ is $\mathcal{F}_t$-measurable,

iii. $\mathbb{P}(\int_0^T f(s, \omega)^2 ds < +\infty) = 1$.

Since the third condition in definition 2.31 is weaker than the third condition in the definition 2.28, we have $\mathcal{V} \subset \mathcal{W}$. We can now define the Itô integral of a process in $\mathcal{W}$. The only difference with the Itô integral of a process in $\mathcal{V}$ as defined previously is that now, all convergence properties are in probability, not in norm $L^2(\Omega)$, which is the 2-th mean convergence that we defined in definition 2.25. Therefore, convergence is in a weaker sense when dealing with processes in $\mathcal{W}$.

Definition 2.32. (Itô integral of a process in $\mathcal{W}$) Let $f \in \mathcal{W}$. Let $(h_n)_{n \in \mathbb{N}}$ be a sequence of elementary processes of $\mathcal{W}$ which converges in probability to f. Then the *Itô integral* of f over $[0, T]$ is defined as:

$$\int_0^T f \mathrm{d}B_t = \lim_{n \to +\infty} \int_0^T h_n \mathrm{d}B_t, \tag{2.34}$$

where the limit is taken in probability.

The existence of the sequence $(h_n)_{n \in \mathbb{N}}$ in definition 2.32 is ensured by a proposition similar to proposition 2.4 with $\mathcal{V}$ replaced by $\mathcal{W}$ and convergence in the norm $\|.\|_{L^2(B)}^2$ (*i.e.* 2-th mean convergence) replaced by convergence in probability. Uniform convergence for $t \in [0, T]$ still holds.

In conclusion, we defined the Itô integral of a process over any interval $[0, T]$ as a limit of a sequence of random variables, either taken in the 2-th mean convergence sense or in the convergence in probability sense. We now extend the definition of the Itô integral of a stochastic process to that of a random linear operator (see [59] chapter 6).

Definition 2.33. (Itô integral of a random linear operator) Let $A : [0, T] \times \Omega \to L(\mathbb{R}^k, \mathbb{R}^n)$ be a random linear operator (for fixed (t, ω), $A(t, \omega)$ is an $n \times k$ matrix). Let $B_t = \begin{bmatrix} B_t^{(1)} & B_t^{(2)} & \dots & B_t^{(k)} \end{bmatrix}^\top$ be a k-dimensional brownian motion. Let $t_i^{(n)} = i\frac{T}{n}$ such that $t_0^{(n)} < t_1^{(n)} < \cdots < t_n^{(n)}$ is a subdivision of $[0, T]$. The *Itô integral of the random linear operator A* with respect to $(B_t)_t$ is defined as the following limit:

$$\int_0^T A(t) \mathrm{d}B_t = \lim_{n \to +\infty} \sum_{i=0}^{n-1} A(t_i^{(n)}) \left(B_{t_{i+1}^{(n)}} - B_{t_i^{(n)}} \right). \tag{2.35}$$

We must make a few remarks to enlighten definition 2.33. Please note that this definition makes sense only if the limit exists, and if we specify which kind of limit we take. Also, if it exists, $\int_0^T A(t) dB_t$ is an n-dimensional random vector. According to [59], if A is adapted with respect to the filtration generated by $(B_t)_t$, and if the $n \times k$ coefficients A_i^j of A verify:

$$\mathbb{P} \left(\int_0^T A_i^j(s, \omega)^2 ds < +\infty \right) = 1, \tag{2.36}$$

then the limit exists in $L^2(\Omega)$, and *a fortiori* exists in probability.

50

2.2.2 Stochastic differential equations

Now that we have defined Itô's integral of a stochastic process, we are in a position to define stochastic differential equations, which are at the heart of this thesis. We are using them to model the dynamics of the sea surface reflectivity.

Definition 2.34. (Itô process) Let $(X_t)_t$ be a n-dimensional stochastic process. We say that $(X_t)_t$ is an *Itô process* if there exists $(a_t)_t$ a n-dimensional stochastic process, $(A_t)_t$ a random linear operator and $(B_t)_t$ a k-dimensional brownian motion such that for all t:

$$X_t = X_0 + \int_0^t a_s \mathrm{d}s + \int_0^t A_s \mathrm{d}B_s, \tag{2.37}$$

with $X_0 = X_{t=0}$. We require that the trajectories of $(a_t)_t$ are almost surely of finite variation over compact of $\mathbb{R}^+$ and that $(A_t)_t$ is adapted to the filtration generated by $(B_t)_t$ and satisfies equation (2.36).

Note that the term $\int_0^t a_s \mathrm{d}s$ is just the time integral of $(a_t)_t$ where integration is carried trajectory wise.

Definition 2.35. (Stochastic differential of an Itô process) Let $(X_t)_t$ be a n-dimensional Itô process such that:

$$X_t = X_0 + \int_0^t a_s \mathrm{d}s + \int_0^t A_s \mathrm{d}B_s. \tag{2.38}$$

We rewrite equation (2.38):

$$\mathrm{d}X_t = a_t \mathrm{d}t + A_t \mathrm{d}B_t. \tag{2.39}$$

$\mathrm{d}X_t$ is called the *stochastic differential* of X_t.

It is manifest in definition 2.35 that the differential notation of a stochastic process is just an other way of writing the integral equation (2.38). We now give two important properties of stochastic differentials.

Proposition 2.5. *(Itô product rule)* *Let X and Y be two one-dimensional Itô processes such that:*

$$\begin{cases} \mathrm{d}X_t = a_t^{(1)}\mathrm{d}t + b_t^{(1)}\mathrm{d}B_t, \\ \mathrm{d}Y_t = a_t^{(2)}\mathrm{d}t + b_t^{(2)}\mathrm{d}B_t. \end{cases} \tag{2.40}$$

Then XY has the following differential:

$$\mathrm{d}(X_t Y_t) = X_t \mathrm{d}Y_t + Y_t \mathrm{d}X_t + b_t^{(1)} b_t^{(2)} \mathrm{d}t, \tag{2.41}$$

which can be written in integral notation:

$$X_t Y_t = \int_0^t X_s a_s^{(2)} ds + \int_0^t X_s b_s^{(2)} dB_s + \int_0^t Y_s a_s^{(1)} ds + \int_0^t Y_s b_s^{(1)} dB_s + \int_0^t b_s^{(1)} b_s^{(2)} \mathrm{d}s. \tag{2.42}$$

The reader should know that actually for any two semimartingales X, Y:

$$\mathrm{d}(X_t Y_t) = X_t \mathrm{d}Y_t + Y_t \mathrm{d}X_t + \mathrm{d}\langle X, Y \rangle_t, \tag{2.43}$$

where $\langle X, Y \rangle_t$ is the quadratic covariation of X and Y. We will introduce the notion of quadratic variation later, in section 2.4 which is about volatility estimation. Of course, for Itô processes verifying equation (2.40), equation (2.43) reduces to equation (2.41).

Proposition 2.6. (Itô formula) *Let* $X = \begin{bmatrix} X^{(1)} & X^{(2)} & \dots & X^{(n)} \end{bmatrix}^\top$ *be a n-dimensional Itô process such that:*

$$\mathrm{d}X_t = a_t \mathrm{d}t + A_t \mathrm{d}B_t, \tag{2.44}$$

where B_t is a k-dimensional brownian motion, and $f : (\mathbb{R}^+ \times \mathbb{R}^n) \to \mathbb{R}^p$ a C^2 application. We denote $f = \begin{bmatrix} f_1 & f_2 & \dots & f_p \end{bmatrix}^\top$. The process $Y_t = f(t, X_t)$ is a p-dimensional process. It holds that the u-th component $Y_t^{(u)}$ verifies:

$$\mathrm{d}Y_t^{(u)} = \frac{\partial f_u}{\partial t}(t, X_t)\mathrm{d}t + \sum_{i=1}^{n} \frac{\partial f_u}{\partial x_i}(t, X_t)\mathrm{d}X_t^{(i)} + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f_u}{\partial x_i \partial x_j}(t, X_t) \sum_{l=1}^{k} A_i^l(t) A_j^l(t)\mathrm{d}t. \tag{2.45}$$

$A_i^l(t)$ is the coefficient of A_t at the i-th row and l-th column. It is manifest that equation (2.45) is component wise for f, such that we can consider that f takes values in $\mathbb{R}$ without loss of generality. Formula (2.45) can then be stated in a somewhat more compact and general form, using again the quadratic variation. Indeed, let $f : (\mathbb{R}^+ \times \mathbb{R}^m) \to \mathbb{R}$ a C^2 function. Let $Y_t = f(t, X_t)$. Then:

$$\mathrm{d}Y_t = \frac{\partial f}{\partial t}(t, X_t)\mathrm{d}t + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(t, X_t)\mathrm{d}X_t^{(i)} + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(t, X_t)\mathrm{d}\langle X^{(i)}, X^{(j)}\rangle_t. \tag{2.46}$$

It is immediate that these two formula are equivalent with:

$$\mathrm{d}\langle X^{(i)}, X^{(j)}\rangle_t = \sum_{l=1}^{k} A_i^l(t) A_j^l(t)\mathrm{d}t. \tag{2.47}$$

Please note that Itô formula is applicable for applications which take values in $\mathbb{C}^p$.

Finally, we define stochastic differential equations.

Definition 2.36. (Stochastic differential equation) Let $a : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^n$ and $A : \mathbb{R}^+ \times \mathbb{R}^n \to L(\mathbb{R}^k, \mathbb{R}^n)$ be two measurable applications (with respect to the Borel σ-algebras). Let $B = \begin{bmatrix} B^{(1)} & B^{(2)} & \dots & B^{(k)} \end{bmatrix}^\top$ be a k-dimensional brownian motion. Let Z be a n-dimensional random vector. A *stochastic differential equation* is an equation of the following form:

$$\begin{cases} \mathrm{d}X_t = a(t, X_t)\mathrm{d}t + A(t, X_t)\mathrm{d}B_t, \\ X_0 = Z, \end{cases} \tag{2.48}$$

which in integral form becomes:

$$X_t = Z + \int_0^t a(s, X_s)\mathrm{d}s + \int_0^t A(s, X_s)\mathrm{d}B_s. \tag{2.49}$$

We say that the stochastic process X_t (n-dimensional) is a *diffusion process*.

a is called *drift* and is often denoted μ instead of a. A is the *volatility* and is often denoted σ or Σ instead of A. In this thesis, we need to work only with *homogeneous* stochastic differential equations, for which the drift and volatility do not depend explicitely on time. If X_t is solution to such a stochastic differential equation, we write:

$$\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t. \tag{2.50}$$

We often leave the initial condition unspecified. Moreover, for homogeneous stochastic differential equations, we may impose the initial condition at any time $t_0 \geq 0$. It is not obvious that a stochastic differential equation like equation (2.50) has a solution. There are theorems specially dedicated to that matter and answering the question: "under which conditions does a stochastic differential equation have a solution, and only one?". We do not need to trouble about that here. We redirect the reader to [105] and [86] (or any textbook on stochastic differential equations). Amongst the conditions in these theorems, we usually require that the initial condition Z be independent of the σ-algebra $\sigma(B_s, s > t_0)$ if the initial condition is imposed at t_0. Also, to have unicity, we require that the solution is adapted to the σ-algebra generated by the initial condition Z and the brownian motion. From now on, we will always assume that there is a unique solution to all the stochastic differential equations we work with. Also we now use the acronym SDE in place of stochastic differential equation.

Theorem 2.1. (*Markov property of a diffusion process*) *Let $(X_t)_t$ be the solution to the SDE (2.48). Then it holds that $(X_t)_t$ is a Markov process. If the SDE is homogeneous, then it holds that $(X_t)_t$ is a homogeneous Markov process (see definition 2.12).*

Theorem 2.1 is of the utmost importance. Since we will encounter only homogeneous SDE, our diffusion processes will always be homogeneous Markov processes. We demonstrate in section 2.3 why this is such a useful property for parameter estimation.

Theorem 2.2. (*semimartingale property of a diffusion process*) *Let $(X_t)_t$ be the solution to the SDE (2.48). Then it holds that $(X_t)_t$ is semimartingale (definition 2.21). More precisely,*

$$A_t = X_0 + \int_0^t \mu(X_s)\mathrm{d}s \tag{2.51}$$

is a finite variation process and

$$M_t = \int_0^t \sigma(X_s)\mathrm{d}s \tag{2.52}$$

is a martingale.

Proposition 2.7. (*Expectation of a diffusion process*) *Let $\int_0^t A(s)\mathrm{d}B_s$ be the Itô integral of a random linear operator with respect to a k-dimensional brownian motion (see definition 2.33). Then:*

$$\mathbb{E}\left[\int_0^t A(s)\mathrm{d}B_s\right] = 0. \tag{2.53}$$

In particular, let $(X_t)_t$ be a diffusion process such that:

$$X_t = X_0 + \int_0^t \mu(s, X_s)\mathrm{d}s + \int_0^t \sigma(s, X_s)\mathrm{d}B_s. \tag{2.54}$$

Then:

$$\mathbb{E}[X_t] = \mathbb{E}[X_0] + \mathbb{E}\left[\int_0^t \mu(s, X_s)ds\right]. \tag{2.55}$$

Proposition 2.7 means that the expectation of a diffusion process is related to the drift μ of its SDE, but not its volatility σ.

2.2.3   Numerical schemes

In the introduction of this chapter, we advised readers uninterested in the mathematics behind stochastic differential equations (SDE) to jump directly here to understand what they are. A SDE can be written:

$$\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t, \tag{2.56}$$

where B_t is a brownian motion. μ is called the drift, and σ the volatility. If we set $\sigma = 0$, equation (2.56) becomes:

$$\mathrm{d}X_t = \mu(X_t)\mathrm{d}t, \tag{2.57}$$

i.e.

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = \mu(X_t), \tag{2.58}$$

which is just an ordinary differential equation. In that case, Euler's scheme states that the increment of X_t over an interval Δt can be approximated by:

$$\Delta X_t \approx \mu(X_t)\Delta t, \tag{2.59}$$

where $\Delta X_t = X_{t+\Delta t} - X_t$. If the initial condition X_0 is deterministic, there is nothing random in equation (2.57) and its solution is a unique trajectory. Now for a SDE, the increment over Δt can be approximated by the Euler's Maruyama scheme which states that:

$$\Delta X_t \approx \mu(X_t)\Delta t + \sigma(X_t)\Delta B_t, \tag{2.60}$$

where $\Delta B_t \sim \mathcal{N}(0, \Delta t)$. To the deterministic increments $\mu(X_t)\Delta t$, we now add a random increment which is normally distributed, with mean zero and variance Δt. Even with a deterministic initial condition, X_t is now a random process due to the random increments: there are infinitely many possible trajectories. In terms of vocabulary, the solution of a SDE is called a *diffusion process*.

Solving analytically a SDE is in most cases impossible. To be more explicit, to solve equation (2.56), one must define a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and find the solution stochastic process defined on Ω. It means that one must determine the trajectories $X_t(\omega)$ for all $\omega \in \Omega$. To remediate to our inability to solve SDE analytically, we use numerical schemes. We refer the reader to [62] for a gentle introduction to numerical resolution of SDE, and to [82] for a more advanced presentation. We now use nD in place of n-dimensional. We define here two different numerical schemes: Euler-Maruyama scheme in the general case of a multidimensional SDE (1D included), and Milstein scheme for 1D processes. Euler-Maruyama's scheme, defined now, is just the repetition of the approximation in equation (2.60) for many intervals.

Definition 2.37. (Euler-Maruyama scheme) Let $[0, T]$ be a finite time interval and $t_0 = 0 < t_1 < \cdots < t_N = T$ a subdivision of $[0, T]$. Let X_t be a multidimensional diffusion process. The Euler-Maruyama scheme reads:

$$X_{t_i} = X_{t_{i-1}} + \mu(X_{t_{i-1}})(t_i - t_{i-1}) + \sigma(X_{t_{i-1}})(B_{t_i} - B_{t_{i-1}}). \tag{2.61}$$

By definition of a brownian motion, $B_{t_i} - B_{t_{i-1}}$ is a vector of independent and centered Gaussian distributed random variable, of variance $\Delta t = t_i - t_{i-1}$. If $X_{t_{i-1}}$ is fixed (deterministic), Euler-Maruyama scheme states that:

$$X_{t_i} - X_{t_{i-1}} \sim \mathcal{N}(\Delta t \mu(X_{t_{i-1}}), \Delta t \sigma(X_{t_{i-1}})\sigma(X_{t_{i-1}})^\top), \tag{2.62}$$

i.e. the increment $X_{t_i} - X_{t_{i-1}}$ is a multivariate Gaussian variable with expectation $\Delta t \mu(X_{t_{i-1}})$ and covariance matrix $\Delta t \sigma(X_{t_{i-1}})\sigma(X_{t_{i-1}})^\top$.

How would Euler-Maruyama go in practice? First assume $t_i - t_{i-1} = \Delta t$ for all i. To generate one trajectory (which is, somehow, an approximation to the trajectory $X_t(\omega)$ for some ω), we generate a starting value according to the distribution of X_{t_0}, and we generate a series of N n-dimensional brownian increments $B_{t_i} - B_{t_{i-1}}$ with normal distribution $\mathcal{N}(0, \Delta t I)$. We use equation (2.61) with these increments to iteratively compute X_{t_i} for all i. To generate many trajectories and numerically see the statistical behaviour of the process, one must repeat this procedure many times to generate many trajectories. Applying this scheme to the SDE:

$$\mathrm{d}X_t = \mathrm{d}B_t, \tag{2.63}$$

with the initial condition $X_0 = 0$, we obtain the trajectories of the brownian motion B_t, which are represented in figure 2.1. We see that all the possible realizations of the process start at $X_0 = 0$ and spread out at t increases. Similar behaviour holds for all diffusion processes and explains why we talk about *diffusions*: the trajectories diffuse away from the initial condition as t increases (for a deterministic initial condition).
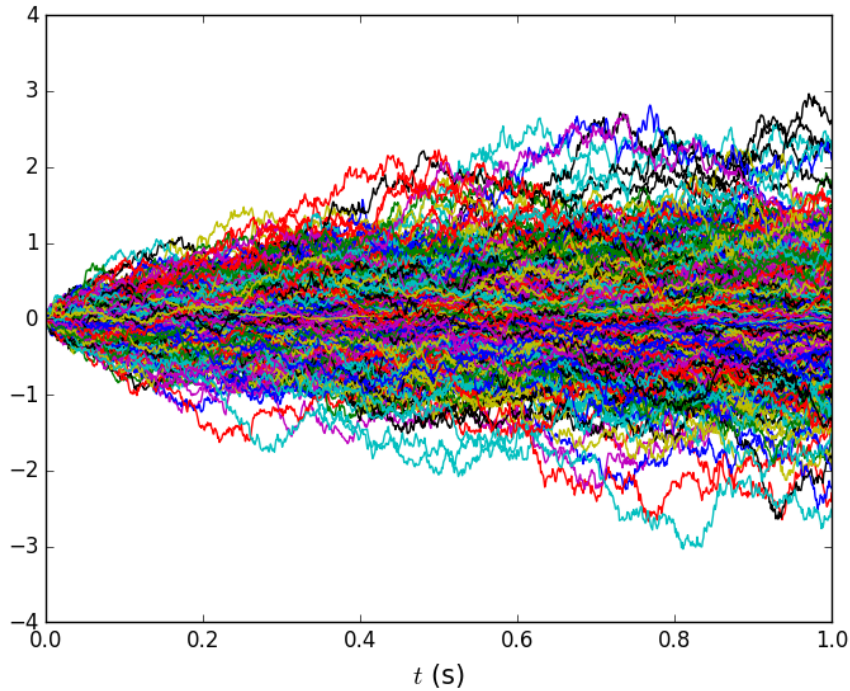
Figure 2.1: 1000 trajectories of a brownian motion generated by Euler-Maruyama's scheme with $\Delta t = 10^{-3}$ s over the time interval $[0, 1]$ s.

Definition 2.38. (Milstein scheme) Let $[0, T]$ be a finite time interval and $t_0 = 0 < t_1 < ... < t_N$ a partition of $[0, T]$. Let X_t be a one-dimensional diffusion process, solution to the SDE (2.56). The Milstein scheme reads:

$$\begin{aligned} X_{t_i} &= X_{t_{i-1}} + \mu(X_{t_{i-1}})(t_i - t_{i-1}) + \sigma(X_{t_{i-1}})(B_{t_i} - B_{t_{i-1}}) \\ &+ \frac{1}{2}\sigma(X_{t_{i-1}})\frac{d\sigma}{dx}(X_{t_{i-1}})((B_{t_i} - B_{t_{i-1}})^2 - \Delta t). \end{aligned} \tag{2.64}$$

55

Milstein scheme is Euler-Maruyama scheme with a correction term which provides it with better convergence properties than Euler-Maruyama scheme. Indeed, it has strong order of convergence 1, while Euler-Maruyama scheme has strong order of convergence 1/2 [62]. Its definition is easy only for 1D SDE and becomes much more tricky for dimensions strictly greater than 1. The simulation of trajectories with Milstein scheme is straightforward, as with Euler-Maruyama scheme.

We will use Euler-Maruyama's scheme to simulate numerical trajectories everytime this is possible. However, Milstein scheme will be necessary to simulate trajectories of the process denoted x_t in the subsequent chapters.

2.3   Transition Probabilities

From now on, we will write X_t for a process $(X_t)_t$. It is an annoying habbit in the literature and we decide to adopt it because it lightens the notations.

2.3.1   Maximum-Likelihood parameter estimation

The maximum-likelihood parameter estimation is classicaly used in signal processing and machine learning and need not be introduced (see *e.g.* [26]). However, it is almost systematically presented for independent random variables. For example, let $Y_1, Y_2, \ldots, Y_n$ be independent and identically distributed random variables according to some probability density $f_\lambda(y)$ parameterized by λ. By independence, the likelihood of an observation $\tilde{Y}_1, \tilde{Y}_2, \ldots, \tilde{Y}_n$ is simply:

$$\mathcal{L}(\tilde{Y}_1, \tilde{Y}_2, \ldots, \tilde{Y}_n; \lambda) = \prod_{i=1}^{n} f_\lambda(\tilde{Y}_i). \tag{2.65}$$

If the Y_i are not independent, it is not true anymore that the likelihood is given by equation (2.65). If the Y_i correspond to the X_{t_i} of a diffusion process with constant timestep $t_i - t_{i-1} = \Delta t$, and if Δt is small, independence is false. Using the Markov property of such diffusion processes, we show now that the likelihood is essentially a product of *transition probabilities* (equation (2.68)), which replace the stationary probabilities $f_\lambda(\tilde{Y}_k)$ of equation (2.65). Thus, for maximum likelihood parameter estimation of diffusion processes, the important concept is that of transition probabilities.

Let us consider a diffusion process X_t, solution to a SDE:

$$\mathrm{d}X_t^{(\lambda)} = \mu_\lambda(X_t^{(\lambda)})\mathrm{d}t + \sigma_\lambda(X_t^{(\lambda)})\mathrm{d}W_t. \tag{2.66}$$

$X_t^{(\lambda)}$ is a n-dimensional stochastic process, μ is a n-dimensional vector, σ is a $n \times k$ matrix (operator) and W_t is a k-dimensional brownian process. The initial condition X_0 is left unprecised. More importantly, we consider that there is a parameter $\lambda = \begin{bmatrix} \lambda_1 & \lambda_2 & \ldots & \lambda_l \end{bmatrix}' \in \mathbb{R}^l$ which parameterize the process. Assume that a particular trajectory of X_t is observed at discrete times $t_0 < t_1 < \ldots < t_N$ and yields $\tilde{X}_0, \tilde{X}_1, \ldots, \tilde{X}_N$.

Since X_t is a Markov process (dependence upon the most recent past), its joint probability density function associated is:

$$\begin{aligned} &p_\lambda(X_{t_0} = \tilde{X}_0, \ldots, X_{t_N} = \tilde{X}_N) \\ = \ &p_\lambda(X_{t_0} = \tilde{X}_0, \ldots, X_{t_{N-1}} = \tilde{X}_{N-1})p_\lambda(X_{t_N} = \tilde{X}_N \mid X_{t_0} = \tilde{X}_0, \ldots, X_{t_{N-1}} = \tilde{X}_{N-1}) \\ = \ &p_\lambda(X_{t_0} = \tilde{X}_0, \ldots, X_{t_{N-1}} = \tilde{X}_{N-1})p_\lambda(X_{t_N} = \tilde{X}_N \mid X_{t_{N-1}} = \tilde{X}_{N-1}). \end{aligned} \tag{2.67}$$

By recurrence, the joint probability density function, that we call *likelihood*, is:

$$\mathcal{L}(\lambda) = p_\lambda(X_{t_0} = \tilde{X}_0) \prod_{i=1}^{N} p_\lambda(X_{t_i} = \tilde{X}_i \mid X_{t_{i-1}} = \tilde{X}_{i-1}). \tag{2.68}$$

Improperly speaking, it is the 'probability' that we observe the sequence $\tilde{X}_0, \tilde{X}_1, \ldots, \tilde{X}_N$ if the parameter is λ. Maximum likelihood estimation (ML) consists in maximizing $\mathcal{L}$ with respect to λ, which yields the estimated parameter. The ML estimator is:

$$\tilde{\lambda} = argmax_\lambda \mathcal{L}(\lambda). \tag{2.69}$$

In practice, we express the optimality conditions:

$$\begin{cases} \frac{\partial \mathcal{L}(\lambda)}{\partial \lambda_1}(\tilde{\lambda}) = 0 \\ \frac{\partial \mathcal{L}(\lambda)}{\partial \lambda_2}(\tilde{\lambda}) = 0 \\ \ldots \\ \frac{\partial \mathcal{L}(\lambda)}{\partial \lambda_l}(\tilde{\lambda}) = 0. \end{cases} \tag{2.70}$$

The solutions to this system are extrema of $\mathcal{L}(\lambda)$. It should be noted that if several solutions exist, one should in theory verify which one maximizes the likelihood. This procedure will be applied several times in this thesis, for estimating the sea clutter parameters, but also the target parameters when there is a target (*e.g.* a boat) in addition to the clutter. To express the system of equations (2.70), one must have the expression of the likelihood in equation (2.68), which in turn requires the initial probability $p_\lambda(X_{t_0} = \tilde{X}_0)$ and the transition probabilities $p_\lambda(X_{t_i} = \tilde{X}_i \mid X_{t_{i-1}} = \tilde{X}_{i-1})$. For illustrative purposes, we represent the first values of one trajectory of a diffusion process in figure 2.2.

2.3.2 The Fokker-Planck equation

It is possible to derive an equation, called the Fokker-Planck equation, which solution is $p_\lambda(X_{t_0} = \tilde{X}_0)$ for some initial conditions (stationary distribution) and $p_\lambda(X_{t_i} = \tilde{X}_i \mid X_{t_{i-1}} = \tilde{X}_{i-1})$ for some other initial conditions (dirac distribution). The Fokker-Planck equation is a partial differential equation. We will not prove it rigorously but choose to present it as a special case of a more general equation (the master equation).

We first present the Fokker-Planck equation for 1D processes. Let X_t be a real-valued diffusion process. We show briefly the important steps that lead to the Fokker-Planck equation. For detailed and rigorous presentations, refer to chapter 4 of [56], or to [120]. The first result which holds for Markov process is the *Chapman-Kolmogorov equation*.

Proposition 2.8. (Chapman-Kolmogorov equation) *Let $t_1 < t_2 < t_3$ and X_t be a Markov process. It then holds:*

$$p(X_{t_3} = x_3 \mid X_{t_1} = x_1) = \int_{\mathbb{R}} p(X_{t_3} = x_3 \mid X_{t_2} = x_2)p(X_{t_2} = x_2 \mid X_{t_1} = x_1)\mathrm{d}x_2. \tag{2.71}$$

The next step is to get the master equation. To do so, we developp the transition probability for short time increments in the following way:

$$p(X_t = x \mid X_{t_0} = x_0) \approx \delta(x - x_0)[1 - a^{(0)}(x_0, t)\Delta t] + W_t(x|x_0)\Delta t, \tag{2.72}$$

with $\Delta t = t - t_0$. $W_t(x|x_0)$ is a transition probability from x_0 to x per unit time. The coefficient $1 - a^{(0)}(x_0, t)\Delta t$ corresponds to the probability that no transition occurs (*i.e.* $x = x_0$).
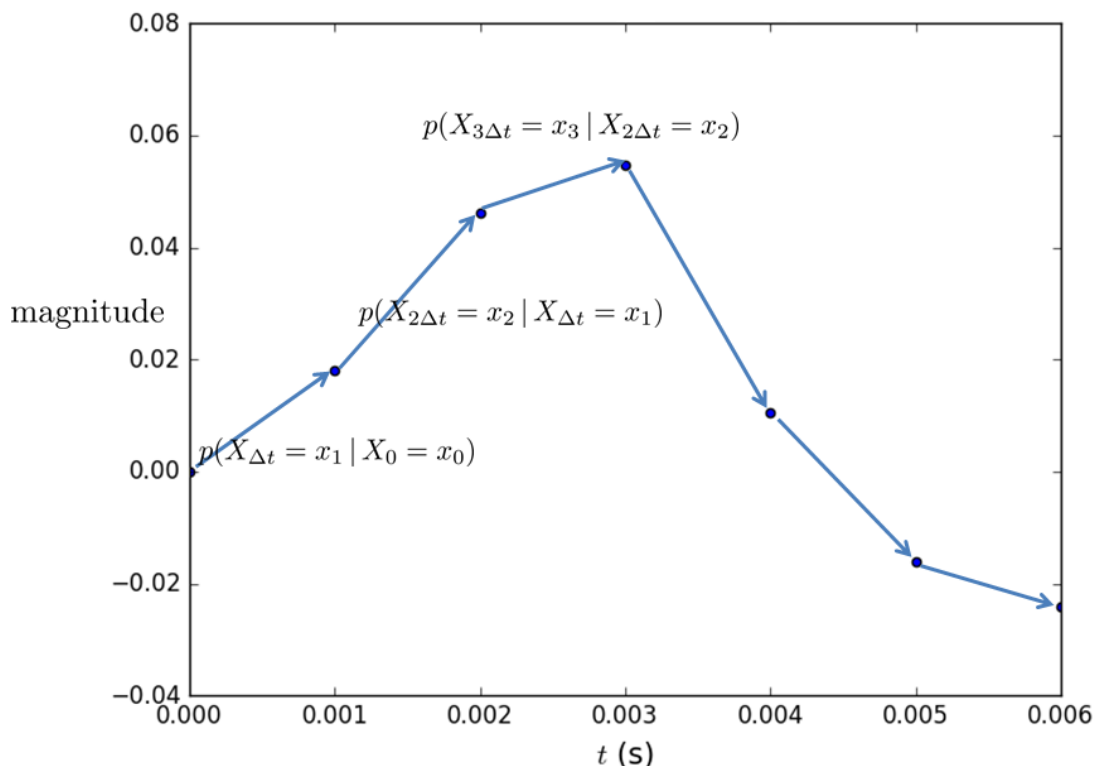
Figure 2.2: First transitions of a diffusion process on the time interval $]0, 0.006]$ s with timestep 0.001 s.

Proposition 2.9. *(master equation) Let X_t be a Markov process. Let t_0, x_0 be fixed, and $p(x,t) = p(X_t = x \mid X_0 = x_0)$. Assume that the transition probability per unit time $W_t(x|x_0)$ is time-homogeneous i.e. $W_t(x|x_0) = W(x|x_0)$. Then it holds:*

$$\frac{\partial p(x,t)}{\partial t} = \int_{\mathbb{R}} [W(x|x')p(x',t) - W(x'|x)p(x,t)]\mathrm{d}x'. \tag{2.73}$$

The next step is *Kramers-Moyal expansion* of the master equation.

Proposition 2.10. *(Kramers-Moyal expansion) Let X_t be a Markov process. Starting from its master equation, one can show that:*

$$\frac{\partial p(x,t)}{\partial t} = \sum_{m=1}^{+\infty} \frac{(-1)^m}{m!} \frac{\partial^m}{\partial x^m} [a^{(m)}(x,t)p(x,t)], \tag{2.74}$$

with $a^{(m)}(x,t) = \int_{\mathbb{R}} r^m W(x;r)\mathrm{d}r$ and $W(x;r) = W(x+r|x)$.

Finally, there is a theorem called *Pawula theorem* (see [120] section 4.3), which states that one cannot truncate the Kramers-Moyal expansion in an arbitrary way. To ensure positive transition probabilities $p(x,t)$ one must either keep only the first term, the first two terms, or the whole series. For homogeneous diffusion processes, only the first two terms remain (the others vanish exactly, such that no approximation is made), which yields the *Fokker-Planck equation*.

Theorem 2.3. (*1D Fokker-Planck equation*) *Let X_t be a 1D homogeneous diffusion process, solution to the SDE:*

$$\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t, \qquad (2.75)$$

Then its density function $(x,t) \mapsto p(X_t = x)$ is solution to the following partial differential equation:

$$\frac{\partial p(X_t = x)}{\partial t} = -\frac{\partial}{\partial x}[\mu(x)p(X_t = x)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[\sigma^2(x)p(X_t = x)]. \qquad (2.76)$$

Equation (2.76) is Kramers-Moyal expansion of the master equation of X_t, truncated after the first two terms. The Fokker-Planck equation is also called *Kolmogorov forward equation*. To solve this equation, one must define an initial condition $x \mapsto p(X_0 = x)$, which is the distribution of the initial condition X_0 of the SDE.

Theorem 2.4. (*nD Fokker-Planck equation*) *Let X_t be a nD homogeneous diffusion process, solution to the SDE:*

$$\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t. \qquad (2.77)$$

$\mu = \begin{bmatrix} \mu_1 & \mu_2 & \ldots & \mu_n \end{bmatrix}^\top$ *is a vector valued application on $\mathbb{R}^n$, and*

$$\sigma = \begin{bmatrix} \sigma_1^1 & \sigma_1^2 & \ldots & \sigma_1^k \\ \sigma_2^1 & \ddots & \ldots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_n^1 & \ldots & \ldots & \sigma_n^k \end{bmatrix} \qquad (2.78)$$

is a $n \times k$ matrix-valued application on $\mathbb{R}^n$, and $W_t = \begin{bmatrix} W_t^{(1)} & W_t^{(2)} & \ldots & W_t^{(k)} \end{bmatrix}^\top$ is a k-dimensional brownian motion. Let $(x,t) \mapsto p(X_t = x)$ for $x = \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix}^\top \in \mathbb{R}^n$, be the distribution of X_t. Then the n-dimensional Fokker-Planck equation is the following partial differential equation:

$$\frac{\partial p(x,t)}{\partial t} = -\sum_{i=1}^n \frac{\partial}{\partial x_i}\left[\mu_i(x,t)p(x,t)\right] + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \frac{\partial^2}{\partial x_i \partial x_j}\left[\Sigma_{ij}(x,t)p(x,t)\right], \qquad (2.79)$$

with $\Sigma = \sigma\sigma^\top$ ($n \times n$ matrix-valued application on $\mathbb{R}^n$).

As for the 1D case, equation (2.79), if solved, gives the distribution $x \mapsto p(X_t = x)$ of X_t for all t. However, to solve it, one must specify an initial condition, *i.e.* $x \mapsto p(X_0 = x)$. We give two very important special cases now.

Proposition 2.11. *Deterministic initial condition Let X_t be a nD homogeneous diffusion process. Assume that $X_0 = x_0$ where $x_0 \in \mathbb{R}^n$. The initial condition for the corresponding Fokker-Planck equation is the distribution δ_{x_0}. It then holds that:*

$$p(X_t = x) = p(X_t = x \mid X_0 = x_0). \qquad (2.80)$$

This result seems obvious, but it is truly a result that the distribution of X_t which starts from a deterministic (Dirac distributed) condition x_0, corresponds to the transition probability $p(X_t = x \mid X_0 = x_0)$ (see [55] p 171).

Proposition 2.12. *Stationary initial condition* *Let X_t be a nD homogeneous diffusion process. The stationary (no time dependence) Fokker-Planck equation is:*

$$0 = -\sum_{i=1}^{n} \frac{\partial}{\partial x_i} \left[\mu_i(x,t)p(x,t)\right] + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2}{\partial x_i\, \partial x_j} \left[\Sigma_{ij}(x,t)p(x,t)\right]. \tag{2.81}$$

Let us denote $x \mapsto p_\infty(x)$ the solution of equation (2.81), which we call stationary distribution. Now assume that X_0 is distributed according to the stationary solution: $x \mapsto p(X_0 = x) = p_\infty(x)$. Then it holds that $\forall t$:

$$p(X_t = x) = p_\infty(x). \tag{2.82}$$

Proposition 2.12 says that if the process X_t is initially distributed according to the stationary distribution, it stays so for all subsequent times. The stationary distribution is also called *asymptotic distribution*. The rationale is that for homogeneous diffusion processes X_t like we consider, whatever the initial condition X_0, the distribution of X_t converges to $p_\infty(x)$ as $t \to +\infty$ (see [120] section 6.1). This result justifies that we use the symbol '∞'. It is not true that all diffusion processes have stationary distribution. For example, the simplest of all:

$$\mathrm{d}X_t = \mathrm{d}B_t \tag{2.83}$$

does not! Indeed, its solution for the initial condition $X_0 = 0$ is $X_t = B_t \sim \mathcal{N}(0,t)$, which does not converge as $t \to +\infty$. Another way to see it is to notice that its stationary FPE:

$$0 = \frac{1}{2}\frac{\partial^2}{\partial x^2}p_\infty(x) \tag{2.84}$$

implies that $p_\infty(x) = ax + b$, which cannot hold since it is required that p_∞ is normalized (integral equal to one).

One reason why we introduced the Fokker-Planck equation is for the purposes of section 2.3.1. Indeed, for maximum likelihood (ML) parameter estimation in the case of Markov processes, we needed to know $p_\lambda(X_{t_0} = \tilde{X}_0)$ and the transition probabilities $p_\lambda(X_{t_i} = \tilde{X}_i \mid X_{t_{i-1}} = \tilde{X}_{i-1})$ (see equation (2.68)). The stationary solution can often be found analytically, but the Fokker-Planck equation with a Dirac initial condition (or any initial condition other than the stationary one) is usually unsolvable analytically. We address this issue in the next section. That being said, we solve analytically and illustrate with simulations the Fokker-Planck equation of two different diffusion processes in chapter 4. In this chapter, transition probabilities are used for probabilistic inference and not ML estimation. They are used for ML estimation in chapters 5 and 6. From now on, we abbreviate Fokker-Planck equation by FPE.

2.3.3   Approximate transition probabilities

How do we get transition probabilities if we cannot solve analytically the FPE? We answer this question in this section because in practice, it is very unlikely that we manage to solve the FPE. A first solution would be to solve it numerically. Though literature exists on the subject ([137], [147], [146], [148], [84], [111]), we consider another simpler option which works well for parameter estimation. If the sampling frequency is high (small Δt), we can approximate very easily the transition probabilities using Euler-Maruyama's scheme.

Consider a diffusion process X_t solution of the SDE:

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t. \tag{2.85}$$

For small Δt, one can advocate Euler-Maruyama scheme (section 2.2.3):

$$X_{t_i} = X_{t_{i-1}} + \mu(X_{t_{i-1}})(t_i - t_{i-1}) + \sigma(X_{t_{i-1}})(W_{t_i} - W_{t_{i-1}}). \tag{2.86}$$

We remind that this scheme is defined in the general case of a n-dimensional SDE driven by a k-dimensional brownian motion. Assume that $X_{t_{i-1}}$ is fixed at the deterministic value x_{i-1}. Then equation (2.86) becomes:

$$X_{t_i} = x_{i-1} + \mu(x_{i-1})(t_i - t_{i-1}) + \sigma(x_{i-1})(W_{t_i} - W_{t_{i-1}}). \tag{2.87}$$

The only stochastic term in the right-hand side of equation (2.87) is $W_{t_i} - W_{t_{i-1}}$, which is a k-dimensional random vector with distribution $\mathcal{N}(0, \Delta t I)$ (I is the identity matrix) from the properties of brownian motion. It is then immediate from equation (2.87) that under the condition $X_{t_{i-1}} = x_{i-1}$, X_{t_i} has a multivariate Gaussian distribution $\mathcal{N}(x_{i-1} + \mu(x_{i-1})\Delta t, \Delta t \sigma(x_{i-1})\sigma(x_{i-1})^\top)$ with $\Delta t = t_i - t_{i-1}$. The approximate transition probability using Euler-Maruyama scheme is:

$$p\left(X_{t_i} = x \mid X_{t_{i-1}} = y\right) \approx \frac{1}{(2\pi)^{n/2} \left|\Delta t \sigma(y)\sigma(y)^\top\right|^{1/2}} \times$$
$$\exp\left(-\frac{1}{2}(x - (y + \mu(y)\Delta t))^\top \left(\Delta t \sigma(y)\sigma(y)^\top\right)^{-1} (x - (y + \mu(y)\Delta t))\right). \tag{2.88}$$

The advantage of equation (2.88) is that it is particularly simple since the transition probability is always Gaussian. Analytical minimization of the likelihood function (in the context of ML estimation) becomes possible (see chapter 5 and 6). The disadvantage is that it works only for small Δt. In particular, the application of transition probabilities in chapter 4 cannot be reduced to this approximation. There are other ways to approximate the transition probabilities. We shall first cite Milstein scheme, in replacement of Euler-Maruyama's scheme, and Aït-Sahalia analytical approximation, based on the expansion of the transition probabilities in Gram-Charlier series (see [7], [36] chapter 6, and [81] for the Gram-Charlier series).

2.4 Volatility Estimation

We are now going to introduce the last important notions that will be required mostly in chapter 7, where volatility estimation is used to estimate parameters. We have seen that a SDE can be written:

$$dX_t = \mu dt + \Sigma dW_t, \tag{2.89}$$

where μ is the drift and Σ is the volatility. The drift and volatility are not symmetric. Under certain conditions, it is possible to make the drift disappear by a change of probability measure ($\mathbb{P}$ on the space Ω). It is quite delicate to estimate the drift and it requires long trajectories [119]. On contrary, the volatility is somewhat more objective than the drift and can be estimated easily. In this section, we explain how we can do so. Estimating the volatility is useful because it enables in turn to estimate volatility parameters (see chapter 7). To be more precise in the terms, we are going to introduce *integrated* volatility estimation, by opposition to *spot* volatility (see [8]). The first step is to define the covariation of two random processes.

Definition 2.39. (Covariation of two real-valued random processes) Let $\Delta = (t_k)_k$ be a subdivision of $\mathbb{R}^+$ and X_t, Y_t be two real-valued random processes. Let $|\Delta| = \sup_{k \in \mathbb{N}}(t_{k+1} - t_k)$. We call *covariation of X and Y* the following limit (in probability):

$$\langle X, Y \rangle_t = \lim_{|\Delta| \to 0} \sum_{k \in \mathbb{N}} (X_{t_{k+1} \wedge t} - X_{t_k \wedge t})(Y_{t_{k+1} \wedge t} - Y_{t_k \wedge t}), \qquad (2.90)$$

where $t_k \wedge t = \min(t_k, t)$ is used to stop the sum at time t.

One can restrict to a bounded interval $[0, T]$ instead of $\mathbb{R}^+$. In that case we choose the subdivision $t_i^{(n)} = i\frac{T}{n}$, and the covariation is:

$$\langle X, Y \rangle_t = \lim_{n \to +\infty} \sum_{k=0}^{n-1} (X_{t_{k+1}} - X_{t_k})(Y_{t_{k+1}} - Y_{t_k}). \qquad (2.91)$$

The covariation is a bilinear operator which maps two processes onto one process.

Notation: The quadratic covariation of a process X_t with itself is called quadratic variation of X_t and is denoted $\langle X \rangle_t$.

Definition 2.40. (Covariation of two n-dimensional random processes) Let $\Delta = (t_k)_k$ be a subdivision of $\mathbb{R}^+$ and X_t, Y_t be two n-dimensional random processes. Let $|\Delta| = \sup_{k \in \mathbb{N}}(t_{k+1} - t_k)$. We call *covariation of X and Y* the following limit (in probability):

$$\langle X, Y \rangle_t = \lim_{|\Delta| \to 0} \sum_{k \in \mathbb{N}} (X_{t_{k+1} \wedge t} - X_{t_k \wedge t})(Y_{t_{k+1} \wedge t} - Y_{t_k \wedge t})^\top. \qquad (2.92)$$

$\langle X, Y \rangle_t$ in equation (2.92) is a $n \times n$ matrix. If $\langle X, Y \rangle_t^{(i,j)}$ denotes its i-th row, j-th column coefficient, then:

$$\langle X, Y \rangle_t^{(i,j)} = \langle X_i, Y_j \rangle_t, \qquad (2.93)$$

i.e. it is the covariation of the i-th component of X_t and the j-th component of Y_t.

One may wonder under which conditions the covariation between two processes exists. All we need to know here is that it exists for any two semimartingales, and thus for any two diffusion processes. We now prove that the covariation of any process with a finite variation process is always zero.

Proposition 2.13. *(Covariation with a finite variation process)* *Let X_t, A_t be two real-valued stochastic processes such that A_t has finite variation (see definition 2.20). Let V_t be the variation of A_t. Then $\forall t \geq 0, \langle X, A \rangle_t = 0$. Indeed, let Δ be a subdivision of $\mathbb{R}^+$:*

$$\left| \sum_{k \in \mathbb{N}} (X_{t_{k+1} \wedge t} - X_{t_k \wedge t})(A_{t_{k+1} \wedge t} - A_{t_k \wedge t}) \right| \leq \sum_{k \in \mathbb{N}} |X_{t_{k+1} \wedge t} - X_{t_k \wedge t}||A_{t_{k+1} \wedge t} - A_{t_k \wedge t}|$$

$$\leq \sup_{\{k \in \mathbb{N}\}} \{|X_{t_{k+1} \wedge t} - X_{t_k \wedge t}|\} \sum_{k \in \mathbb{N}} |A_{t_{k+1} \wedge t} - A_{t_k \wedge t}|$$

$$\leq \sup_{\{k \in \mathbb{N}\}} \{|X_{t_{k+1} \wedge t} - X_{t_k \wedge t}|\} V_t \xrightarrow{|\Delta| \to 0} 0. \qquad (2.94)$$

That $\sup_{\{k \in \mathbb{N}\}} \{|X_{t_{k+1} \wedge t} - X_{t_k \wedge t}|\} \xrightarrow{|\Delta| \to 0} 0$ is not obvious. However, it is true if we assume that X_t is a cadlàg process (trajectories are continuous from the right, and admit a limit from the left). To justify that, see [12] proposition 5.

Proposition 2.14. *(Covariation of brownian motions)* *Let $W_t^{(1)}$, $W_t^{(2)}$ be two independent brownian motions. Then it holds:*

    i. $\langle W^{(i)}, W^{(i)} \rangle_t = t$ *for $i = 1, 2$,*

    ii. $\langle W^{(1)}, W^{(2)} \rangle_t = 0$.

Proposition 2.15. *(Covariation of integrated processes)* *Let X_t, Y_t be two real-valued processes which are the Itô integrals of the processes $\sigma_t^{(1)}, \sigma_t^{(2)}$ respectively:*

$$X_t = \int_0^t \sigma_s^{(1)} \mathrm{d}W_s^{(1)}, \tag{2.95}$$

$$Y_t = \int_0^t \sigma_s^{(2)} \mathrm{d}W_s^{(2)}. \tag{2.96}$$

Then we have:

$$\langle X, Y \rangle_t = \int_0^t \sigma_s^{(1)} \sigma_s^{(2)} d\langle W^{(1)}, W^{(2)} \rangle_s. \tag{2.97}$$

For us, $\langle W^{(1)}, W^{(2)} \rangle_t$ is always either t (if $W_t^{(1)} = W_t^{(2)}$) or 0 (if $W_t^{(1)}$ and $W_t^{(2)}$ are independent). We have now all the results to state the main results to compute covariations in the framework of diffusion processes.

Proposition 2.16. *Let $W_t^{(j)}$ for $j = 1, 2, \ldots, k$ be k independent brownian motions. Let X_t and Y_t be two Itô processes whose stochastic differentials are:*

$$\begin{cases} \mathrm{d}X_t = \mu^{(X)} \mathrm{d}t + \sum_{j=1}^k \sigma_j^{(X)} \mathrm{d}W_t^{(j)} \\ \mathrm{d}Y_t = \mu^{(Y)} \mathrm{d}t + \sum_{j=1}^k \sigma_j^{(Y)} \mathrm{d}W_t^{(j)}. \end{cases} \tag{2.98}$$

Then by the bilinearity of the quadratic covariation and by proposition 2.15, we have:

$$\mathrm{d}\langle X, Y \rangle_t = \sum_{j=1}^k \sigma_j^{(X)} \sigma_j^{(Y)} \mathrm{d}t. \tag{2.99}$$

Equation (2.99) is obtained by the previous results but it can be obtained from equation (2.98) simply by a couple of rules of calculus. Indeed, we simply have:

$$\mathrm{d}\langle X, Y \rangle_t = \mathrm{d}X_t \mathrm{d}Y_t, \tag{2.100}$$

with the two following rules:

1. a product with $\mathrm{d}t$ gives 0,

2. a product of two brownian increments $\mathrm{d}W_t^{(\alpha)}$ and $\mathrm{d}W_t^{(\beta)}$ gives $\mathrm{d}t$ if $W_t^{(\alpha)} = W_t^{(\beta)}$, 0 otherwise.

It is assumed here that two brownian motions are either identical, or independent. Equation (2.100) is *the* most important result of this section to compute quadratic covariation in practice (see chapter 7).

Proposition 2.17. *Let $X = \begin{bmatrix} X^{(1)} & X^{(2)} \ldots & X^{(n)} \end{bmatrix}^\top$ be a n-dimensional Itô process with stochastic differential:*

$$\mathrm{d}X_t = \mu\mathrm{d}t + \Sigma\mathrm{d}W_t, \tag{2.101}$$

where $W_t = \begin{bmatrix} W_t^{(1)} & W_t^{(2)} & \ldots & W_t^{(k)} \end{bmatrix}^\top$ is a k-dimensional brownian motion. Σ is a $n \times k$ matrix. Σ_i^l denotes its i-th row and l-th column coefficient.

Then, the quadratic variation of X verifies:

$$\mathrm{d}\langle X\rangle_t = \Sigma\Sigma^\top\mathrm{d}t = \sigma\mathrm{d}t. \tag{2.102}$$

It is a $n \times n$ matrix whose i-th row, j-th column coefficient is:

$$\sigma^{ij}\mathrm{d}t = \mathrm{d}\langle X^{(i)}, X^{(j)}\rangle_t = \sum_{l=1}^{k}\Sigma_i^l\Sigma_j^l\mathrm{d}t. \tag{2.103}$$

Note that in propositions 2.16 and 2.17, the processes are not necessarily diffusion processes, *i.e.* solutions to SDE. They are just Itô processes.

Finally, volatility-based parameter estimation goes as follows. Let Y_t be a n-dimensional process solution of the stochastic differential equation (SDE):

$$\mathrm{d}Y_t = \mu(Y_t)\mathrm{d}t + \Sigma(Y_t)\mathrm{d}W_t. \tag{2.104}$$

Let $\sigma = \Sigma\Sigma^\top$. If σ^{ij} is the i-th row, j-th column coefficient of σ and $Y^{(k)}$ the k-th coordinate of Y, then:

$$\langle Y^{(i)}, Y^{(j)}\rangle_t = \int_0^t \sigma^{ij}(Y_s)\mathrm{d}s. \tag{2.105}$$

$\langle Y^{(i)}, Y^{(j)}\rangle_t$ is the quadratic covariation of $Y^{(i)}$ and $Y^{(j)}$. Let $t_k^{(n)} = k\frac{t}{n}$ be a subdivision of $[0, t]$ into n pieces. Then by definition:

$$\langle Y^{(i)}, Y^{(j)}\rangle_t = \lim_{n\to+\infty}\sum_{k=0}^{n-1}(Y_{t_{k+1}}^{(i)} - Y_{t_k}^{(i)})(Y_{t_{k+1}}^{(j)} - Y_{t_k}^{(j)}). \tag{2.106}$$

It holds from equations (2.105) and (2.106) that:

$$\sum_{k=0}^{n-1}\sigma^{ij}(Y_{t_k})\Delta t \approx \sum_{k=0}^{n-1}(Y_{t_{k+1}}^{(i)} - Y_{t_k}^{(i)})(Y_{t_{k+1}}^{(j)} - Y_{t_k}^{(j)}), \tag{2.107}$$

with $\Delta t = \frac{t}{n}$. Equation (2.107) is the cornerstone of volatility-based parameter estimation. If $\sigma^{ij}(Y_{t_k})$ depends on some parameters to be estimated and the $Y_t^{(k)}$ are observed, we can isolate the parameters in terms of known quantities and get estimators.

We wish to distinguish between two terms: volatility estimation and volatility-based estimation. In short, volatility estimation consists in writing equation (2.107). The integral of the squared-volatility is estimated by the sum in the right-hand of equation (2.107). To be more specific, this is integrated volatility estimation, by opposition to spot volatility (see [8], and chapter 7 section 7.4.1). In both cases, it is the volatility itself which is estimated. Besides volatility estimation (integrated or spot), we call volatility-based estimation the estimation of parameters relying on volatility estimation.

2.5 Relation between maximum likelihood and volatility

Volatility-based estimation works for high sampling frequencies (small Δt). Under this hypothesis, the drift increment of a SDE can be neglected because it has order Δt while the volatility increment has order $\Delta t^{1/2}$. We want to show that maximum likelihood and volatility-based estimation have some connection. We think that volatility-based estimation is a simplified version of ML estimation when the parameter to be estimated is in the volatility. Let us consider again the SDE:

$$\mathrm{d}Y_t = \mu(Y_t)\mathrm{d}t + \Sigma_\lambda(Y_t)\mathrm{d}W_t, \tag{2.108}$$

which is now 1D for simplicity. The squared-volatility $\sigma_\lambda = \Sigma_\lambda^2$ depends on a parameter λ. Equation (2.107) becomes:

$$\sum_{k=0}^{n-1} \sigma_\lambda(Y_{t_k})\Delta t - \sum_{k=0}^{n-1}(Y_{t_{k+1}} - Y_{t_k})^2 = 0. \tag{2.109}$$

Solving equation (2.109) for λ gives the volatility-based estimation of λ. For ML, we apply Euler-Maruyama's scheme to (2.108) and neglect the drift increment because of high sampling frequency assumption. We get:

$$Y_{t_{k+1}} - Y_{t_k} = \Sigma_\lambda(Y_{t_k})\Delta W_{t_k}, \tag{2.110}$$

where $\Delta W_{t_k} \sim \mathcal{N}(0, \Delta t)$. We get:

$$Y_{t_{k+1}} \sim \mathcal{N}(Y_{t_k}, \sigma_\lambda(Y_{t_k})\Delta t). \tag{2.111}$$

Therefore, the transition probability from Y_{t_k} to $Y_{t_{k+1}}$ reads:

$$p(Y_{t_{k+1}} \mid Y_{t_k}) = \frac{1}{\sqrt{2\pi\Delta t\sigma_\lambda(Y_{t_k})}}e^{-\frac{1}{2}\frac{\left(Y_{t_{k+1}}-Y_{t_k}\right)^2}{\Delta t\sigma_\lambda(Y_k)}}, \tag{2.112}$$

and the likelihood of the time series $Y = \{Y_0, Y_1, \ldots, Y_n\}$ can be approximated by:

$$\mathcal{L}(Y, \lambda) = \prod_{k=0}^{n-1} \frac{1}{\sqrt{2\pi\Delta t\sigma_\lambda(Y_{t_k})}}e^{-\frac{1}{2}\frac{\left(Y_{t_{k+1}}-Y_{t_k}\right)^2}{\Delta t\sigma_\lambda(Y_k)}}. \tag{2.113}$$

Taking the natural logarithm, we get the log-likelihood:

$$l(Y, \lambda) = -\frac{1}{2}n\ln(2\pi\Delta) - \frac{1}{2}\sum_{k=0}^{n-1}\ln(\sigma_\lambda(Y_k)) - \frac{1}{2}\sum_{k=0}^{n-1}\frac{\left(Y_{t_{k+1}} - Y_{t_k}\right)^2}{\Delta t\sigma_\lambda(Y_k)}. \tag{2.114}$$

To obtain the maximum likelihood estimator, we must maximize $l(Y, \lambda)$ with respect to λ. To do so, we express the optimality condition:

$$\frac{\partial l}{\partial \lambda}(Y, \lambda) = 0. \tag{2.115}$$

If we express this condition from equation (2.114) and multiply it by $-2\Delta t$, we get:

$$\sum_{k=0}^{n-1}\frac{\sigma_\lambda'(Y_{t_k})}{\sigma_\lambda(Y_{t_k})}\Delta t - \sum_{k=0}^{n-1}\left(Y_{t_{k+1}} - Y_{t_k}\right)^2 \frac{\sigma_\lambda'(Y_{t_k})}{\sigma_\lambda^2(Y_{t_k})} = 0, \tag{2.116}$$

where $\sigma'_\lambda = \frac{\partial \sigma_\lambda}{\partial \lambda}$. The λ which solves equation (2.116) is the maximum likelihood estimator of λ, using Euler-Maruyama's scheme to approximate the transition probabilities and neglecting the drift. We see a strong resemblance between equation (2.116) and equation (2.109), which attests to the link between the volatility-based and maximum likelihood estimators. Equations (2.116) and (2.109) are not identical, but they become so if σ_λ does not depend on the state Y_t. It seems that volatility-based estimation is a dumbed down version of maximum likelihood.

2.6   Wiener-Khinchin theorem

In this section, we introduce quickly the Wiener-Khinchin theorem since it is a classical tool for relating the spectral content of a stationary signal (realization of a stationary random process) to its correlation properties.

Let $x : t \mapsto x(t)$ be a deterministic stationary signal. The autocorrelation function of x is the even function defined as:

$$G(\tau) = \lim_{T \to +\infty} \frac{1}{T} \int_0^T x(t)x(t+\tau)\mathrm{d}t, \qquad (2.117)$$

for $\tau \in \mathbb{R}$. We can typically think of $x(t)$ as one realization of a stochastic process. For a pulsation $\omega \in \mathbb{R}$ (not to be mistaken with an element of the probability space Ω), we set:

$$\hat{x}_T(\omega) = \int_0^T x(t)e^{-i\omega t}\mathrm{d}t, \qquad (2.118)$$

and

$$S(\omega) = \frac{1}{2\pi} \lim_{T \to +\infty} \frac{1}{T}|\hat{x}_T(\omega)|^2. \qquad (2.119)$$

This is the definition of the *power spectral density* (PSD) of the signal $x(t)$.

Theorem 2.5. (*Wiener-Khinchin*) *For a signal $x(t)$ with autocorrelation function $G(\tau)$ and PSD $S(\omega)$, it holds:*

$$S(\omega) = \frac{1}{2\pi} \int_\mathbb{R} G(\tau)e^{-i\omega\tau}\mathrm{d}\tau, \qquad (2.120)$$

and

$$G(\tau) = \int_\mathbb{R} S(\omega)e^{i\omega\tau}\mathrm{d}\omega. \qquad (2.121)$$

The Wiener-Khinchin theorem states that the autocorrelation function and the PSD are equivalent under the Fourier transform. Let us now consider a stochastic process X_t. If $\tilde{X}_t = x(t)$ is one realization of the process, we can apply the Wiener-Khinchin theorem to it. Under the assumption that X_t is stationary, we define its autocorrelation function as:

$$G(\tau) = \langle X_t X_{t+\tau} \rangle. \qquad (2.122)$$

If we further assume that it is ergodic, then:

$$G(\tau) = \langle X_t X_{t+\tau} \rangle = \lim_{T \to +\infty} \frac{1}{T} \int_0^T \tilde{X}_t \tilde{X}_{t+\tau}\mathrm{d}t := \tilde{G}(\tau), \qquad (2.123)$$

which makes the connection between the definitions of G for a deterministic and stochastic signal (equations (2.117) and (2.122)). It is then visible that $\lim_{T \to +\infty} \frac{1}{T} \int_0^T \tilde{X}_t \tilde{X}_{t+\tau} \mathrm{d}t$ does not depend on the particular realization of the process. The PSD of the arbitrary realization $\tilde{X}_t$ is denote $\tilde{S}(\omega)$. If we apply the Wiener-Kinchin theorem, we get:

$$\tilde{S}(\omega) = \frac{1}{2\pi} \int_{\mathbb{R}} \tilde{G}(\tau) e^{-i\omega\tau} \mathrm{d}\tau = \frac{1}{2\pi} \int_{\mathbb{R}} G(\tau) e^{-i\omega\tau} \mathrm{d}\tau = S(\omega). \tag{2.124}$$

The last equality holds because since for any realization $\tilde{X}_t$ we have $\tilde{G}(\tau) = G(\tau)$, the PSD of $\tilde{X}_t$ ends up being independent of the particular trajectory.

To sum up, if we have a process X_t whose autocorrelation function $G(\tau) = \langle X_t X_{t+\tau} \rangle$ is known, the PSD $S(\omega)$ is independent of the trajectory and it holds:

$$S(\omega) = \frac{1}{2\pi} \int_{\mathbb{R}} \langle X_t X_{t+\tau} \rangle e^{-i\omega\tau} \mathrm{d}\tau, \tag{2.125}$$

and

$$\langle X_t X_{t+\tau} \rangle = \int_{\mathbb{R}} S(\omega) e^{i\omega\tau} \mathrm{d}\omega. \tag{2.126}$$

In this thesis, the Wiener-Khinchin theorem is only used in chapter 5, section 5.5, in an attempt to define a spectrum-based estimator for one of the parameters of the stochastic model for the sea clutter (introduced in chapter 3).

2.7   Conclusion

In this chapter, we placed ourselves in a purely mathematical framework and introduced many notions with the degree of rigour that seemed necessary to us for a good understanding of the rest of the thesis. The brownian motion was introduced in section 2.1 and we understood its special place amongst continuous stochastic processes, its main property being that it has independent Gaussian increments. It served as a base to define stochastic differential equations in section 2.2, since SDE are driven by a brownian motion (or a martingale more generally). Stochastic processes which are solutions to SDE are called diffusion processes and we have seen that they are both martingales and Markov processes. We saw in section 2.3 that their transition probabilities are solutions of the Fokker-Planck equation, but can be approximated for small Δt using Euler-Maruyama's scheme. Thanks to the Markov property of diffusion processes, transition probabilities can be used to compute the likelihood of a time series, in conjunction with the asymptotic distribution of the process, and carry out parameter estimation. We introduced quadratic covariations in section 2.4. We saw that it relates to the volatility of diffusion processes, and that volatility estimation can be used for parameter estimation. We argued in section 2.5 that volatility-based estimation is a simplified version of maximum likelihood estimation. Finally, Wiener-Khintcin theorem was presented in section 2.6.

In the next chapter, we introduce Field's model, expressed in terms of SDE, for the electromagnetic field scattered by a random medium.

Chapter 3

Field's model for the sea clutter

In chapter 1, section 1.2.6, we defined the sea clutter as being the noise-like signal that we get from illuminating the sea surface with a radar (see figure 1.11). It was not precised whether the radar is in monostatic or bistatic configuration, what polarization is used etc. We assume now that a radar is indeed emitting electromagnetic waves onto a patch of the sea surface, and that this one scatters back part of the waves to a receiver located possibly anywhere. It includes both the monostatic and bistatic configurations, and any polarization (in particular any of the HH, VV, HV, VH polarizations).

At the end of chapter 1, we presented the random walk model for the sea clutter. We explained that the random walk model (section 1.3) does not account for the dynamics of the clutter. It expresses the reflectivity as:

$$\Psi_t = \lim_{\bar{N} \to +\infty} \sum_{n=1}^{N_t} \frac{a_t^{(n)}}{\bar{N}^{1/2}} e^{i\phi_t^{(n)}} = \lim_{\bar{N} \to +\infty} \left(\frac{N_t}{\bar{N}}\right)^{1/2} \lim_{\bar{N} \to +\infty} \sum_{n=1}^{N_t} \frac{a_t^{(n)}}{N_t^{1/2}} e^{i\phi_t^{(n)}} = x_t^{1/2} \gamma_t, \qquad (3.1)$$

where $\bar{N}$ is the average number of scatterers, and N_t the actual number of scatterers at time t. That t appears can be misleading: t is "frozen" in equation (3.1), it is a non-explicited variable. In equation (3.1), we have purposely factorized Ψ_t in two factors:

$$x_t^{1/2} = \lim_{\bar{N} \to +\infty} \left(\frac{N_t}{\bar{N}}\right)^{1/2}, \qquad (3.2)$$

and

$$\gamma_t = \lim_{\bar{N} \to +\infty} \sum_{n=1}^{N_t} \frac{a_t^{(n)}}{N_t^{1/2}} e^{i\phi_t^{(n)}}. \qquad (3.3)$$

x_t (radar cross section) and γ_t (speckle) are solutions to stochastic differential equations (SDE) in Field's model. They are presented in detail respectively in sections 3.1 and 3.2. The factorization of the sea surface reflectivity has been used before, such as in [57], where x_t is called "texture" and does not correspond to a population model, but is a sum of cosines (cyclostationarity) to account for the cyclic motion of the sea surface. In [57], the second factor, γ_t, is also called speckle and has the same statistics as in Field's model, at the exception that it does not solve a SDE.

We explained that if N_t follows a negative binomial distribution and the phases are uniform over $[0, 2\pi[$, $z_t = |\Psi_t|^2$ follows the K distribution. In chapter 2, we introduced mathematical notions and most particularly we defined *stochastic differential equations* (SDE). A

stochastic process X_t may therefore be determined by a SDE:

$$\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t, \tag{3.4}$$

where W_t is a brownian motion. It almost goes without saying that *a SDE dictates the dynamics of its solution*. This can be easily intuited thanks to Euler-Maruyama scheme for example (section 2.2.3): an increment over a small interval Δt is a Gaussian random variable with mean $\mu(X_t)\Delta t$ and variance $\sigma(X_t)^2\Delta t$, or the transition probability over a small time interval is approximately Gaussian (section 2.3.3). Expressing Ψ_t as the solution of a SDE would therefore solve our problem of modelling the *dynamics* of the sea clutter. This is precisely the work done by Field in [48] and presented here.

The presentation of Field's model in [48] is not always very clear, in terms of vocabulary, notations, notions, and proofs. In this chapter, we propose a more detailed presentation of the model. It is already part of the actual work done during the thesis because it goes beyond pure bibliography, unlike chapters 1 and 2.

In section 3.1, we focus on the process for the number of scatterers, N_t, and show how Field shows that its asymptotic normalized counterpart, $x_t = \lim_{\bar{N}\to+\infty} \frac{N_t}{N}$, is solution of a SDE. In section 3.2, we derive the SDE for the reflectivity when the number of scatterers is constant but goes to infinity ($N_t = N \to +\infty$), *i.e.* $\lim_{\bar{N}\to+\infty} \sum_{n=1}^{N_t} \frac{a_t^{(n)}}{N_t^{1/2}} e^{i\phi_t^{(n)}}$. Based on equation (3.1), these two terms constitute the random walk model augmented with SDE for the dynamics, which is exactly Field's model. In section 3.3, we show that Field's model is a generalization of the K distribution, and finally in section 3.4 we summarize the chapter.

3.1 The stochastic population of scatterers

If we analyse equation (3.1), we see essentially three stochastic processes whose dynamics should or could be specified via a SDE: N_t, $a_t^{(n)}$ and $\phi_t^{(n)}$. This section is dedicated to N_t: the number of scatterers contributing to the reflectivity at time t. In the case of the sea surface, we will not study who the scatterers are. They may be crests of waves, or points which are geometrically well located for an optical reflection of the incident waves. The existence of a population of scatterers (identified as small localized areas of the sea surface contributing strongly to the backscattered signal) is somewhat acknowledged in [141] as early in the book as chapter 2 (before the random walk model). There is empirical evidence (see p 31 of [141]) that the scatterers are different for different polarizations: HH, VV, HV, VH.

A crucial hypothesis in what follows is that $\bar{N} \to +\infty$: we assume that the number of scatterers contributing to the sum is very large. In the context of radar remote sensing of the sea surface, this hypothesis is questionable when the illuminated area becomes small (high resolution radars). Mathematically, it is interesting to consider the limit $\bar{N} \to +\infty$ to ensure the continuity of Ψ_t. It is a necessary condition to obtain a model of stochastic differential equations driven by brownian motions. If this hypothesis is removed, Ψ_t becomes a jump process and Field's model [48] as described in this chapter breaks down.

3.1.1 The linear Birth-Death-Immigration model

In Field's model, N_t is a $\mathbb{N}$-valued stochastic process: it is a single population model. On the general subject of population models, we highly recommend [98]. Field, in [48] chapter 7, uses a linear Birth-Death-Immigration (BDI) population model (see [98] chapter 5 for the

theory). Let $P_N(t) = \mathbb{P}(N_t = N)$ be the probability that the population is N at time t. The linear BDI model states that between t and $t + \Delta t$, where Δt is small, there is:

1. either a birth with probability $\lambda \Delta t N$

2. or a death with probability $\mu \Delta t N$

3. or an immigration with probability $\nu \Delta t$

4. or nothing happens with probability $1 - \Delta t \left(\nu + (\lambda + \mu) N \right)$.

The birth and death rates (proportional to N) can be understood quite easily. Immigration may be more delicate to interpret, since it is a constant positive rate independent of the population N. From a mathematical point of view, it is necessary to have immigration to make sure that there exists an equilibrium. In a Birth-Death population model, $N = 0$ is an absorbing state and given some initial condition it is almost sure that N reaches zero (extinction). Adding immigration enables the population to increase again when $N = 0$ is reached [98]. The four possible transitions of the population over a time interval Δt (birth, death, immigration or nothing) are summed up in figure 3.1. If a transition time step h is considered such that $h \gg \Delta t$, the number of the scatterers between t and $t + h$ can change a lot due to multiple births, deaths and immigration, as illustrated in figure 3.2. For more clarity, the scatterers are represented as static. The motion is taken into account when the speckle is studied in section 3.2.
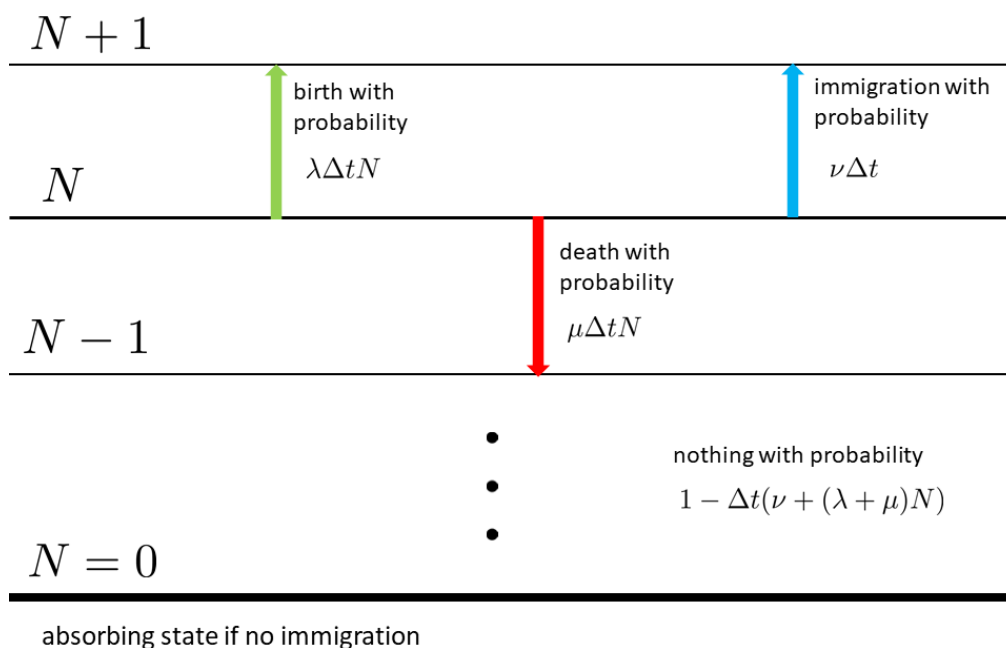
Figure 3.1: Possible transitions of the population N over a time interval Δt.

Then the following equation holds:

$$
\begin{aligned}
P_N(t + \Delta t) &= P_N(t) \left[1 - \Delta t \left(\nu + (\lambda + \mu) N \right) \right] \\
&+ P_{N-1}(t) \Delta t (\lambda(N-1) + \nu) + P_{N+1}(t) \Delta t \mu (N+1),
\end{aligned} \tag{3.5}
$$

71
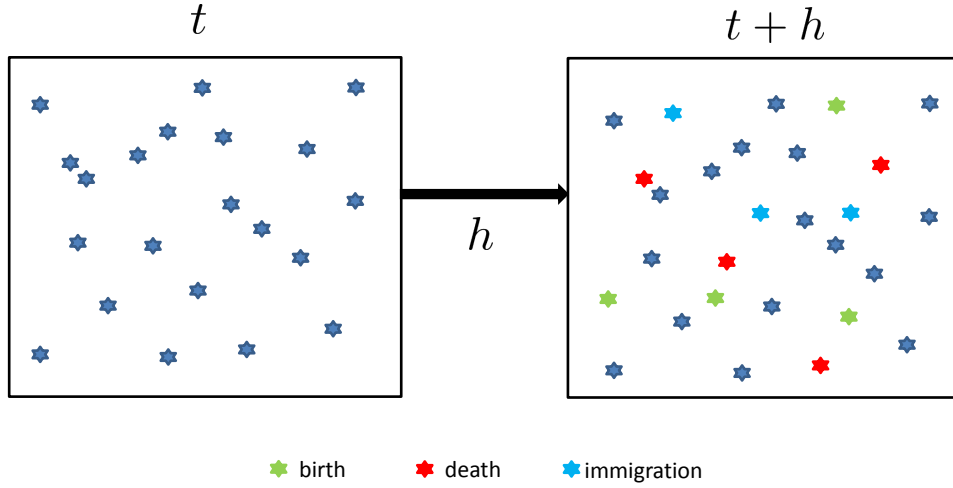
Figure 3.2: Example of a time change of a random population following the BDI model.

which is equivalent to

$$\frac{P_N(t + \Delta t) - P_N(t)}{\Delta t} = (\nu + \lambda(N-1))P_{N-1}(t) - (\nu + (\lambda + \mu)N)P_N(t) + \mu(N+1)P_{N+1}(t). \quad (3.6)$$

Taking the limit for $\Delta t \to 0$, we get:

$$\frac{dP_N(t)}{dt} = (\nu + \lambda(N - 1))P_{N-1}(t) - (\nu + (\lambda + \mu)N)P_N(t) + \mu(N + 1)P_{N+1}(t). \quad (3.7)$$

Equation (3.7) corresponds to equation (7.1) in [48]:

$$\frac{dP_N(t)}{dt} = G_{N-1}P_{N-1}(t) - (G_N + R_N)P_N(t) + R_{N+1}P_{N+1}(t), \quad (3.8)$$

with:

$$\begin{cases} G_N = \nu + \lambda N \\ R_N = \mu N. \end{cases} \quad (3.9)$$

λ is the birth rate, μ is the death rate and ν is the immigration rate. G_N and R_N are respectively the generation and recombination rates.

If N_t is stationary, then $\frac{dP_N(t)}{dt} = 0$ and it is shown in [98] section 5.3.2 that in that case, the stationary (or asymptotic or equilibrium) distribution of N_t is:

$$P_N(t) = \binom{N + \alpha - 1}{N} p^\alpha (1 - p)^N, \quad (3.10)$$

with $\alpha = \frac{\nu}{\lambda}$ and $p = \frac{\mu - \lambda}{\mu}$. If we set $\bar{N} = \frac{\nu}{\mu - \lambda}$, we can show that:

$$\begin{cases} p^\alpha = \frac{1}{1 + \bar{N}/\alpha} \\ (1 - p)^N = \frac{\bar{N}/\alpha}{1 + \bar{N}/\alpha}, \end{cases} \quad (3.11)$$

72

from which we get:

$$P_N(t) = \binom{N + \alpha - 1}{N} \frac{\left(\bar{N}/\alpha\right)^N}{\left(1 + \bar{N}/\alpha\right)^{N+\alpha}}, \tag{3.12}$$

which is precisely the negative binomial distribution of section 1.3. $\bar{N} = \mathbb{E}[p(N_t = N)]$ is the average number of scatterers.

3.1.2 The FPE and SDE for x_t

We see from equation (3.1) for Ψ_t that it is not N_t itself which we must study, but the limit of the ratio $\frac{N_t}{\bar{N}}$ for large $\bar{N}$. $\frac{N_t}{\bar{N}}$ is a random variable which can take all values of the form $k\frac{1}{\bar{N}}$ for $k \in \mathbb{N}$. As $\bar{N} \to +\infty$, these possible values become closer and closer, such that $x_t = \lim_{\bar{N} \to +\infty} \frac{N_t}{\bar{N}}$ takes values in $\mathbb{R}^+$ (continuous random variable).

In what follows, we propose a formal presentation following that of [48] and using the Kramers-Moyal expansion (proposition 2.10). If the demonstration is too technical or unclear, the reader can jump directly to equation (3.26).

Equation (3.8) is the starting point to get a SDE for x_t. In the general case where G_N and R_N depend on time, it can be rewritten:

$$\begin{aligned}
\frac{dP(N,t)}{dt} &= -\left[G(N,t)P(N,t) - G(N-1,t)P(N-1,t)\right] \\
&+ \left[R(N+1,t)P(N+1,t) - R(N,t)P(N,t)\right],
\end{aligned} \tag{3.13}$$

with $P(N,t) = P_N(t)$, $G(N,t) = G_N(t)$ (generation rate) and $R(N,t) = R_N(t)$ (recombination rate). Now by Taylor expansion (with respect to N), this can be rewritten:

$$\frac{dP(N,t)}{dt} = \sum_{n=1}^{+\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial N^n}\left[G(N,t)P(N,t)\right] + \sum_{n=1}^{+\infty} \frac{1}{n!} \frac{\partial^n}{\partial N^n}\left[R(N,t)P(N,t)\right]. \tag{3.14}$$

Equation (3.14) may not be rigourous as it is. We have expanded by Taylor series the functions $G(N,t)P(N,t)$ and $R(N,t)P(N,t)$ with respect to N. This hardly makes sense as $P(N,t)$ is defined on $\mathbb{N}$. However, we admit it formally for the presentation since it is implicitly used by Field in [48]. We use it to identify the coefficients of the Kramers-Moyal expansion of $P(N,t)$ as functions of the generation and recombination rates.

We remind that the Kramers-Moyal expansion for $P(N,t)$ reads (see section 2.3.2):

$$\frac{\partial P(N,t)}{\partial t} = \sum_{n=1}^{+\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial N^n}[a^{(n)}(N,t)P(N,t)], \tag{3.15}$$

written

$$\frac{\partial P(N,t)}{\partial t} = \sum_{n=1}^{+\infty} \left(-\frac{\partial}{\partial N}\right)^n [D^{(n)}(N,t)P(N,t)] \tag{3.16}$$

by Field (equation 7.4 in [48]). By identification with equation (3.8), it is immediate that:

$$D^{(n)}(N,t) = \frac{1}{n!}[G(N,t) + (-1)^n R(N,t)]. \tag{3.17}$$

We could have postulated equation (3.17) to avoid the unappropriate Taylor expansion. The important point is that if equation (3.17) holds, what follows is true. Now let $x = \frac{N}{\bar{N}}$. The Kramers-Moyal expansion for $P(x, t) = P(x\bar{N}, t)\bar{N}$ is:

$$\frac{\partial P(x, t)}{\partial t} = \sum_{n=1}^{+\infty} \frac{1}{\bar{N}^{n-1}} \left(-\frac{\partial}{\partial x} \right)^n [D^{(n)}(x\bar{N}, t)P(x, t)]. \tag{3.18}$$

We now go back to the special case where $G(N, t) = \lambda N + \nu$ and $R(N, t) = \mu N$, i.e. the linear BDI population model. The first term ($n = 1$) is:

$$\frac{1}{\bar{N}^0} \left(-\frac{\partial}{\partial x} \right) [D^{(1)}(x\bar{N}, t)P(x, t)] = \left(-\frac{\partial}{\partial x} \right) [(G(x, t) - R(x, t))P(x, t)], \tag{3.19}$$

with

$$\begin{cases} G(x, t) = \lambda x\bar{N} + \nu \\ R(x, t) = \mu x\bar{N}. \end{cases} \tag{3.20}$$

Reminding that $\bar{N} = \frac{\nu}{\mu - \lambda}$, we easily show that:

$$\frac{1}{\bar{N}^0} \left(-\frac{\partial}{\partial x} \right) [D^{(1)}(x\bar{N}, t)P(x, t)] = -\left(\frac{\partial}{\partial x} \right) (\nu(1 - x)P(x, t)). \tag{3.21}$$

This term does not depend on $\bar{N}$. The second term of the Kramers-Moyal expansion is:

$$\frac{1}{\bar{N}} \left(-\frac{\partial}{\partial x} \right) [D^{(2)}(x\bar{N}, t)P(x, t)] = \left(-\frac{\partial}{\partial x} \right) \left(\frac{1}{2}G(x\bar{N}, t) + R(x\bar{N}, t) \right), \tag{3.22}$$

which can be shown to be:

$$\frac{1}{\bar{N}} \left(-\frac{\partial}{\partial x} \right) [D^{(2)}(x\bar{N}, t)]P(x, t) = \frac{1}{2} \left(\frac{\partial}{\partial x} \right)^2 \left(x(\lambda + \nu) + \frac{\nu}{\bar{N}}P(x, t) \right). \tag{3.23}$$

What happens to equation (3.23) when $\bar{N} \to +\infty$? To make $\bar{N}$ go to infinity, we keep ν fixed and let $\mu \to \lambda$ (or $\lambda \to \mu$). Equation (3.23) tends to:

$$\left(\frac{\partial}{\partial x} \right)^2 [\lambda x P(x, t)]. \tag{3.24}$$

All the terms for $n \geq 3$ tend to zero as $\bar{N} \to +\infty$. Finally, as $\bar{N} \to +\infty$, $x_t = \lim_{\bar{N} \to +\infty} \frac{N_t}{\bar{N}}$ has the distribution $p(x_t = x)$ solution of the partial differential equation:

$$\frac{\partial}{\partial t} p(x_t = x) = -\frac{\partial}{\partial x} [\nu(1 - x)p(x_t = x)] + \frac{\partial^2}{\partial x^2} [\lambda x p(x_t = x)]. \tag{3.25}$$

Equation (3.25) is the Fokker-Planck equation for the stochastic process x_t. If we set $\alpha = \nu/\lambda$ and $\mathcal{A} = \nu$, it becomes:

$$\frac{\partial}{\partial t} p(x_t = x) = -\frac{\partial}{\partial x} [\mathcal{A}(1 - x)p(x_t = x)] + \frac{\partial^2}{\partial x^2} \left[\frac{\mathcal{A}}{\alpha} x p(x_t = x) \right]. \tag{3.26}$$

The stationary distribution of the process x_t is the solution to:

$$0 = -\frac{\partial}{\partial x} [\mathcal{A}(1 - x)p(x_t = x)] + \frac{\partial^2}{\partial x^2} \left[\frac{\mathcal{A}}{\alpha} x p(x_t = x) \right]. \tag{3.27}$$

The solution is a gamma-distribution (see [48] p 49):

$$p(x_t = x) = \frac{\alpha(\alpha x)^{\alpha-1} e^{-\alpha x}}{\Gamma(\alpha)} \tag{3.28}$$

for $x \geq 0$, 0 otherwise. We generally do not feature the indicator function to lighten the notations. The reader shall remember that x_t is always ≥ 0.

We presented the Fokker-Planck equation (FPE) in section 2.3.2. The important result is theorem 2.3, which states that to every SDE, one can associate a FPE for the probability density of the solution X_t of the SDE. Reciprocally, knowing the FPE for x_t, we can say that x_t is solution to a SDE whose FPE is given by equation (3.26). Therefore, x_t is solution to the following SDE:

$$dx_t = \mathcal{A}(1 - x_t)dt + \left(2\frac{\mathcal{A}}{\alpha}x_t\right)^{\frac{1}{2}} dW_t^{(x)}, \tag{3.29}$$

for some 1D brownian motion $W_t^{(x)}$. This is the first brick of Field's model for the sea clutter! A necessary step to arrive at equations (3.26) (Fokker-Planck equation for $p(x_t = x)$) and (3.29) (stochastic differential equation driven by a brownian motion for x_t) is the limiting procedure:

$$\lim_{\bar{N} \to +\infty} \frac{N_t}{\bar{N}}. \tag{3.30}$$

Without it, the population could not be quantified as a continous random variable and we would neither have a FPE nor a SDE.

3.1.3 Discussion on the population model

Two main hypotheses subtend the results of section 3.1. The first assumption comes back to the random walk model and we think that it is important to remind it. There is a population model (whatever the model) because we start from the random walk, which postulates a family of discrete scatterers. This is an approximation of the real EM waves/sea surface interaction which should be justified. Under what circumstances can equation (3.1) for the random walk model, be a good approximation to a more rigorous formulation like the Stratton-Chu equations (1.65)? What would be the discrete scatterers?

The second hypothesis is that the population follows a linear BDI model. To relax this assumption, one could start from equations (3.16) and (3.17), setting arbitrary generation and recombination coefficients $G(N, t)$ and $R(N, t)$, and see where the calculations go. Does the Kramers-Moyal expansion reduces to a FPE when $\bar{N} \to +\infty$? If yes what is the associated SDE? However, it is not obvious that one can take arbitrary $G(N, t)$ and $R(N, t)$ because of Pawula theorem. For example, let's consider a nonlinear BDI model as in [98] chapter 7. In the nonlinear BDI model, the birth and death rate are functions of the population N. The birth rate $\lambda(N)$ is:

$$\lambda(N) = \lambda_0 N - \lambda_1 N^{s+1} \tag{3.31}$$

if $N < (\frac{\lambda_0}{\lambda_1})^{1/s}$, 0 otherwise. $(\frac{\lambda_0}{\lambda_1})^{1/s}$ is the N such that $\lambda(N) = 0$, *i.e.* no more births. In absence of immigration, it would the maximum number of scatterers. The death rate $\mu(N)$ is:

$$\mu(N) = \mu_0 N + \mu_1 N^{s+1}. \tag{3.32}$$

s is an integer ≥ 1. The immigration rate, ν, remains the same. [98] chapter 7 shows that such a nonlinear BDI population verifies the following (Kolmogorov) equation:

$$\begin{aligned}
\frac{dP(N,t)}{dt} &= (\nu + \lambda(N-1))\,P(N-1,t) \\
&\quad - (\nu + \lambda(N) + \mu(N))\,P(N,t) + \mu(N+1)P(N+1,t),
\end{aligned} \qquad (3.33)$$

which is equivalent to

$$\frac{dP(N,t)}{dt} = G_{N-1}P(N-1,t) - (G_N + R_N)P(N,t) + R_{N+1}P(N+1,t) \qquad (3.34)$$

with $G_N = \nu + \lambda(N)$ and $R_N = \mu(N)$. The generation and recombination rates are therefore:

$$\begin{cases} G_N = \nu + (\lambda_0 N - \lambda_1 N^{s+1})I_{[0,N_m]} \\ R_N = \mu_0 N + \mu_1 N^{s+1}. \end{cases} \qquad (3.35)$$

I is the indicator function and $N_m = (\frac{\lambda_0}{\lambda_1})^{1/s}$. If we simplify the model by removing the indicator function, it becomes:

$$\begin{cases} G_N = \nu + (\lambda_0 N - \lambda_1 N^{s+1}) \\ R_N = \mu_0 N + \mu_1 N^{s+1}. \end{cases} \qquad (3.36)$$

It is hard to continue analytically for a simple reason: in general, we do not know the expression of $\bar{N}$ as a function of the parameters λ_0, λ_1, μ_0, μ_1 and ν. Scouting out a little bit actually convinces us that there is little hope that the nonlinear BDI leads to an interesting result for x_t. Let $x = \frac{N}{\bar{N}}$. For example, the first term of the Kramers-Moyal expansion of $P(x,t) = P(x\bar{N},t)\bar{N}$ is:

$$\left(-\frac{\partial}{\partial x}\right)[(G(x,t) - R(x,t))P(x,t)] = -\frac{\partial}{\partial x}\left[\nu + (\lambda_0 - \mu_0)\bar{N}x - (\lambda_1 + \mu_1)\bar{N}^{s+1}x^{s+1}\right]. \qquad (3.37)$$

How do we let $\bar{N} \to +\infty$ and make sure that this first term does not explode? It seems necessary that $\lambda_0 - \mu_0 \to 0$ and $\lambda_1 + \mu_1 \to 0$. This is equivalent to $\mu_0 \to \lambda_0$ (or $\lambda_0 \to \mu_0$) and $\lambda_1 \to 0$ and $\mu_1 \to 0$. But if λ_1 and μ_1 go to zero, we just go back to the linear BDI model! And making μ_0 go to λ_0 is simply the way we made $\bar{N}$ go to infinity in the linear BDI case!

Another method to modify the population model consists in directly postulating a SDE for x_t, as suggested in [48] chapter 10. One would then have to make sure that x_t remains positive. This is ensured for x_t in equation (3.29) because at $x_t = 0$, the volatility vanishes and the drift is $\mathcal{A}$, which is stricly positive. Therefore, $x = 0$ is a barrier that x_t never crosses. A way to find such a SDE is to postulate a stationary distribution for x_t different than equation (3.28), find which FPE it is the stationary solution of (which could be hard!) and obtain the corresponding SDE.

In complete generality, there is no certainty that $x_t = \lim_{\bar{N} \to +\infty} \frac{N_t}{\bar{N}}$ is a diffusion process, *i.e.* solution to a SDE (cf. [48] chapter 7 and [68]).

3.2 The speckle

We started the previous section by stating that there are 3 stochastic processes in the random walk model whose dynamics could be specified via a SDE: N_t, $a_t^{(n)}$ and $\phi_t^{(n)}$. N_t or more

precisely $x_t = \lim_{\bar{N} \to +\infty} \frac{N_t}{N}$, was treated and led to the FPE (3.26) and the SDE (3.29). Field considers that $a_t^{(n)} = a = 1$ is constant ([48] chapter 6). This assumption is relaxed in [48] chapter 10 or [47] and in this section, where $a_t^{(n)} = a_n$ depends on the scatterer. Therefore, we never specify dynamics for the amplitudes $a_t^{(n)}$.

Equation (3.1) can be written:

$$\Psi_t = x_t^{1/2} \lim_{\bar{N} \to +\infty} \sum_{n=1}^{N_t} \frac{a_n}{N_t^{1/2}} e^{i\phi_t^{(n)}}. \tag{3.38}$$

We now wish to derive a SDE for:

$$\gamma_t = \lim_{\bar{N} \to +\infty} \sum_{n=1}^{\bar{N}} \frac{a_n}{\bar{N}^{1/2}} e^{i\phi_t^{(n)}}. \tag{3.39}$$

γ_t is called the *speckle*. This is the name chosen to refer to γ_t, but it may not correspond exactly to what is usually referred to as speckle. For example, there is no space variable here since the processes, like γ_t, correspond to the time evolution of one resolution cell. Since:

$$\gamma_t = \lim_{\bar{N} \to +\infty} \sum_{n=1}^{\bar{N}} \frac{a_n}{\bar{N}^{1/2}} e^{i\phi_t^{(n)}} = \lim_{\bar{N} \to +\infty} \sum_{n=1}^{N_t} \frac{a_n}{N_t^{1/2}} e^{i\phi_t^{(n)}}, \tag{3.40}$$

such a SDE would complete the dynamics of Ψ_t, since the SDE of x_t is already known.

The speckle corresponds to a constant population of scatterers ($\bar{N}$, though it is increased to infinity). Its importance lies in the fact that the phases $\phi_t^{(n)}$ depend on time. In the usual random walk model, the dynamics of the phases are not specified (section 1.3). The speckle is the reflectivity for a constant population of dynamic scatterers, as illustrated in figure 3.3.
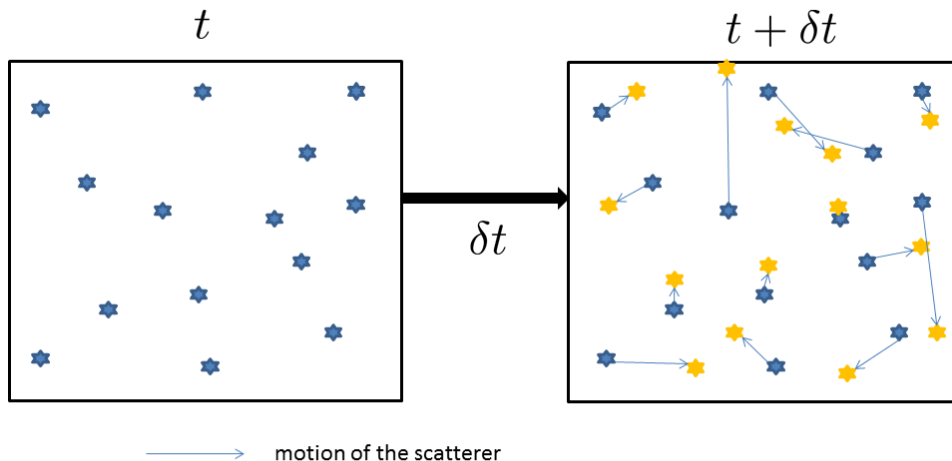
Figure 3.3: Motions of a constant population of scatterers between t and $t + \Delta t$ corresponding to phase changes $\Delta \phi_t^{(n)}$.

In the next section, we present a first model for the phases where the scatterers have independent brownian motion. We account for the situation where a deterministic translational motion affects all the scatterers in section 3.2.2.

3.2.1 Speckle without Doppler

The first case for which we want a SDE for γ_t is the case where the phases are given by the SDE:

$$\begin{cases} \mathrm{d}\phi_t^{(n)} = \mathcal{B}^{1/2}\mathrm{d}W_t^{(n)} \\ \phi_0^{(n)} = \Delta^{(n)}, \end{cases} \tag{3.41}$$

where $\forall n$, $\Delta^{(n)}$ is uniformily distributed over $[0, 2\pi[$ and $\mathcal{B}$ is a positive constant.

Independence is always assumed for the phases: for $i \neq j$, $\phi_t^{(i)}$ and $\phi_t^{(j)}$ are independent. This last assumption is questionable for the sea surface. Indeed, large motions of water masses can transport a group of scatterers in a similar directions, in which case the time evolutions of $\phi_t^{(i)}$ and $\phi_t^{(j)}$ are no longer independent.

Equation (3.41) is the SDE for the unwrapped phase, and it does not have a stationary/asymptotic solution since $\mathcal{B}^{1/2}W_t^{(n)}$ has variance $\mathcal{B}\Delta t$, which goes to $+\infty$ as $t \to +\infty$. However, we understand that the wrapped phase has uniform distribution over $[0, 2\pi[$ for stationary distribution, even though there is no SDE for the wrapped phase since it is a jump process. There is "no Doppler" in the phase model (3.41), in the sense that the scatterers do not have a common component of deterministic translation due to a rectilinear motion of the radar for example (as in satellite or airborne acquisitions). The case where there is such a translational motion is treated in the next section.

To derive the SDE of the speckle γ_t in the no Doppler case, we rely largely on the paper by Feng, Field and Hayking [47] appendices B, C, D. We follow their lead but it would be pointless to give all the details. Instead, we focus on the important steps to understand where the model could be modified. Let

$$\varepsilon_t^{(N)} = \sum_{k=1}^{N} a_k e^{i\phi_t^{(k)}} \tag{3.42}$$

be the finite random walk over N scatterers which have phases $\phi_t^{(1)}, \phi_t^{(2)}, ..., \phi_t^{(N)}$ respectively. We assume that we know the stochastic differentials of the phases. The stochastic differentials will be given by equation (3.41) ultimately, but for now on we stay general. We can compute the differential $\mathrm{d}\varepsilon_t^{(N)}$ by application of Itô formula (proposition 2.6) with:

$$\begin{array}{llll} f: & \mathbb{R}^N & \to & \mathbb{C} \\ (\phi^{(1)}, \phi^{(2)}, ..., \phi^{(N)}) & \mapsto & f(\phi^{(1)}, \phi^{(2)}, ..., \phi^{(N)}) = \sum_{k=1}^{N} a_k e^{i\phi^{(k)}} \end{array} \cdot \tag{3.43}$$

Itô formula gives:

$$\mathrm{d}\varepsilon_t^{(N)} = \sum_{j=1}^{N} \frac{\partial f}{\partial \phi^{(j)}} \mathrm{d}\phi_t^{(j)} + \frac{1}{2} \sum_{j,l=1}^{N} \frac{\partial^2 f}{\partial \phi^{(j)} \partial \phi^{(l)}} \mathrm{d}\langle \phi^{(j)}, \phi^{(l)} \rangle_t, \tag{3.44}$$

since f does not depend explicitly on time. From the expression of f, the second order

derivative $\frac{\partial^2 f}{\partial \phi^{(j)} \partial \phi^{(l)}}$ are zero for $j \neq l$, which yields:

$$
\begin{aligned}
d\varepsilon_t^{(N)} &= \sum_{j=1}^{N} \left[\frac{\partial f}{\partial \phi^{(j)}} d\phi_t^{(j)} + \frac{1}{2} \frac{\partial^2 f}{\partial \phi^{(j)\,2}} d\langle \phi^{(j)}, \phi^{(j)} \rangle_t \right] \\
&= \sum_{j=1}^{N} \left[ia_j e^{i\phi_t^{(j)}} d\phi_t^{(j)} + \frac{1}{2}(-1)a_j e^{i\phi_t^{(j)}} d\langle \phi^{(j)}, \phi^{(j)} \rangle_t \right] \\
&= \sum_{j=1}^{N} a_j \left[id\phi_t^{(j)} - \frac{1}{2} d\langle \phi^{(j)}, \phi^{(j)} \rangle_t \right] e^{i\phi_t^{(j)}}.
\end{aligned}
\tag{3.45}
$$

Now we use the SDE (3.41) for the phases dynamics: we quit generality and enter deep into Field's assumptions. Since $d\langle \phi^{(j)}, \phi^{(j)} \rangle_t = \mathcal{B} dt$, we get:

$$
d\varepsilon_t^{(N)} = \sum_{j=1}^{N} ia_j \mathcal{B}^{1/2} dW_t^{(j)} e^{i\phi_t^{(j)}} - \frac{1}{2} \mathcal{B} dt \sum_{j=1}^{N} a_j e^{i\phi_t^{(j)}}
$$

$$
\Leftrightarrow \quad d\varepsilon_t^{(N)} = -\frac{1}{2} \mathcal{B} \varepsilon_t^{(N)} dt + V
\tag{3.46}
$$

with $V = \sum_{j=1}^{N} ia_j \mathcal{B}^{1/2} dW_t^{(j)} e^{i\phi_t^{(j)}}$ (same notations as in [47]). There is nothing to do with $-\frac{1}{2} \mathcal{B} \varepsilon_t^{(N)} dt$: it is directly interpreted as a part (or the whole) of the drift of a SDE for $d\varepsilon_t^{(N)}$. It is V which causes problem and must be transformed to a more friendly form. V can be written into real and imaginary parts:

$$
V = \mathcal{B}^{1/2} \left[i \sum_{j=1}^{N} a_j \cos(\phi_t^{(j)}) dW_t^{(j)} - \sum_{j=1}^{N} a_j \sin(\phi_t^{(j)}) dW_t^{(j)} \right].
\tag{3.47}
$$

These real and imaginary part (disregarding the constant factor $\mathcal{B}^{1/2}$) can be written:

$$
\begin{aligned}
\sum_{j=1}^{N} a_j \cos(\phi_t^{(j)}) dW_t^{(j)} &= \left(\sum_{j=1}^{N} a_j^2 \cos^2(\phi_t^{(j)}) \right)^{1/2} \frac{\sum_{j=1}^{N} a_j \cos(\phi_t^{(j)}) dW_t^{(j)}}{\left(\sum_{j=1}^{N} a_j^2 \cos^2(\phi_t^{(j)}) \right)^{1/2}} \\
&= \sigma_c dW_t^{(c)},
\end{aligned}
\tag{3.48}
$$

with

$$
\begin{cases}
\sigma_c^2 = \sum_{j=1}^{N} a_j^2 \cos^2(\phi_t^{(j)}) \\
dW_t^{(c)} = \frac{\sum_{j=1}^{N} a_j \cos(\phi_t^{(j)}) dW_t^{(j)}}{\left(\sum_{j=1}^{N} a_j^2 \cos^2(\phi_t^{(j)}) \right)^{1/2}}.
\end{cases}
\tag{3.49}
$$

where $W_t^{(c)}$ is a brownian motion. The main property used here, is that $dW_t^{(c)}$ is indeed the stochastic differential of a brownian motion. This is true by independence of the $W_t^{(j)}$. Similarly, we can show that:

$$
\sum_{j=1}^{N} a_j \sin(\phi_t^{(j)}) dW_t^{(j)} = \sigma_s dW_t^{(s)},
$$

with

$$\begin{cases} \sigma_s^2 = \sum_{j=1}^{N} a_j^2 \sin^2(\phi_t^{(j)}) \\ \mathrm{d}W_t^{(s)} = \frac{\sum_{j=1}^{N} a_j \sin(\phi_t^{(j)}) \mathrm{d}W_t^{(j)}}{\left(\sum_{j=1}^{N} a_j^2 \sin^2(\phi_t^{(j)}) \right)^{1/2}}. \end{cases} \tag{3.50}$$

where $W_t^{(s)}$ is a brownian motion. V can now be rewritten:

$$V = \mathcal{B}^{1/2} \sigma \left[i \frac{\sigma_c}{\sigma} \mathrm{d}W_t^{(c)} - \frac{\sigma_s}{\sigma} \mathrm{d}W_t^{(s)} \right], \tag{3.51}$$

with

$$\sigma^2 = \sigma_c^2 + \sigma_s^2 = \sum_{j=1}^{N} a_j^2. \tag{3.52}$$

In complete generality for N, one cannot go further. However, [47] shows that:

$$\begin{cases} \lim_{N \to +\infty} \langle W_t^{(c)}, W_t^{(s)} \rangle_t = 0 \\ \lim_{N \to +\infty} \left(\frac{\sigma_c}{\sigma} \right)^2 = \lim_{N \to +\infty} \left(\frac{\sigma_s}{\sigma} \right)^2 = \frac{1}{2}. \end{cases} \tag{3.53}$$

From the first equality, he deduces the independence of $W_t^{(c)}$ and $W_t^{(s)}$ for infinite N (using [79]). If the a_j are independent of N, then σ explodes as $N \to +\infty$ (unless we impose some decaying condition on the sequence $\{a_j, j \in \mathbb{N}\}$, which is not relevant). It is also assumed that $\sup_{j \in \mathbb{N}}(a_j) < +\infty$. To prevent the explosion of σ, we normalize the a_j by $N^{1/2}$, which is exactly what we do in equation 3.39 when we define the speckle γ_t. Thus, σ^2 becomes:

$$\sigma^2 = \frac{1}{N} \sum_{j=1}^{N} a_j^2, \tag{3.54}$$

and has a finite limit for $N \to +\infty$. We can deduce that as $N \to +\infty$, V becomes:

$$V = \frac{\mathcal{B}^{1/2} \sigma}{\sqrt{2}} \left[-\mathrm{d}W_t^{(s)} + i\mathrm{d}W_t^{(c)} \right] = \mathcal{B}^{1/2} \sigma \mathrm{d}\xi_t, \tag{3.55}$$

where

$$\xi_t = \frac{1}{\sqrt{2}} (-W_t^{(s)} + iW_t^{(c)}) \tag{3.56}$$

is a complex brownian motion. We remind that:

$$\gamma_t = \lim_{N \to +\infty} \sum_{j=1}^{N} \frac{a_j}{N^{1/2}} e^{i\phi_t^{(j)}} = \lim_{N \to +\infty} \frac{\varepsilon_t^{(N)}}{N^{1/2}}, \tag{3.57}$$

From equations (3.46) and (3.55), we get the following SDE for the speckle:

$$\mathrm{d}\gamma_t = -\frac{1}{2} \mathcal{B} \gamma_t \mathrm{d}t + \mathcal{B}^{1/2} \sigma \mathrm{d}\xi_t. \tag{3.58}$$

This is the second main SDE of Field's model, together with the SDE (3.29) for x_t. Field makes the additional assumption that $\sigma = 1$, Which has the effect of normalizing the reflectivity (see section 3.3). Contrary to Field in [48], we prefer to rewrite equation (3.58) in the form of two real-valued SDE rather than one complex valued SDE. It gives:

$$\begin{cases} \mathrm{d}\gamma_t^{(R)} = -\frac{1}{2} \mathcal{B} \gamma_t^{(R)} \mathrm{d}t + \frac{1}{\sqrt{2}} \mathcal{B}^{\frac{1}{2}} \mathrm{d}W_t^{(R)} \\ \mathrm{d}\gamma_t^{(I)} = -\frac{1}{2} \mathcal{B} \gamma_t^{(I)} \mathrm{d}t + \frac{1}{\sqrt{2}} \mathcal{B}^{\frac{1}{2}} \mathrm{d}W_t^{(I)}. \end{cases} \tag{3.59}$$

$\gamma_t^{(R)}$ and $\gamma_t^{(I)}$ are respectively the real and imaginary parts of γ_t. $W_t^{(R)}$ and $W_t^{(I)}$ are two real-valued independent brownian motion. They are related to ξ_t by:

$$\xi_t = \frac{1}{\sqrt{2}}(W_t^{(R)} + iW_t^{(I)}). \tag{3.60}$$

We see immediately that $W_t^{(R)} = -W_t^{(s)}$ and $W_t^{(I)} = W_t^{(c)}$, hence the independence.

3.2.2 Speckle with Doppler

We presented the Doppler effect in section 1.2.4. We have seen that if the relative velocity between the radar and object is $\vec{V}$, and if $\vec{u}_r$ is unitary and points from the object to the radar, then when an EM wave of frequency f_0 is emitted toward the object, the received frequency is:

$$f_D = f_0 \frac{c_0}{c_0 + 2\vec{V}.\vec{u}_r} = f_0 + \left(\frac{c_0}{c_0 + 2\vec{V}.\vec{u}_r} - 1\right) f_0 = f_0 + f_{doppler}, \tag{3.61}$$

with $f_{doppler} = \left(\frac{c_0}{c_0+2\vec{V}.\vec{u}_r} - 1\right) f_0$. Let $\omega_0 = 2\pi f_{doppler}$. The phase of the reflectivity in baseband (i.e. after multiplication of the received complex signal by $e^{-i2\pi f_0 t}$) is:

$$\phi_t^{(j)} = -\omega_0 t + constant. \tag{3.62}$$

In terms of differentials:

$$\mathrm{d}\phi_t^{(j)} = -\omega_0 \mathrm{d}t. \tag{3.63}$$

Equation (3.63) is for a linearly moving scatterer. If we add a brownian component to the phase dynamics via the SDE (3.41), we obtain:

$$\mathrm{d}\phi_t^{(j)} = -\omega_0 \mathrm{d}t + \mathcal{B}^{1/2}\mathrm{d}W_t^{(j)}. \tag{3.64}$$

We can repeat the calculations of section 3.2.1 very easily and obtain a SDE for γ_t with the new phase model (3.64). Let

$$\varepsilon_t^{(N)} = \sum_{k=1}^{N} a_k e^{i\phi_t^{(k)}}. \tag{3.65}$$

We have shown in complete generality with respect to the phase model that:

$$\mathrm{d}\varepsilon_t^{(N)} = \sum_{j=1}^{N} a_j \left[i\mathrm{d}\phi_t^{(j)} - \frac{1}{2}\mathrm{d}\langle\phi^{(j)}, \phi^{(j)}\rangle_t\right] e^{i\phi_t^{(j)}}. \tag{3.66}$$

Since we still have $\langle\phi^{(j)}, \phi^{(j)}\rangle_t = \mathcal{B}\mathrm{d}t$, we get:

$$
\begin{aligned}
\mathrm{d}\varepsilon_t^{(N)} &= \sum_{j=1}^{N} a_j \left[i\left(-\omega_0\mathrm{d}t + \mathcal{B}^{1/2}\mathrm{d}W_t^{(j)}\right) - \frac{1}{2}\mathcal{B}\mathrm{d}t\right] e^{i\phi_t^{(j)}} \\
&= \sum_{j=1}^{N} a_j e^{i\phi_t^{(j)}} \left(-i\omega_0 - \frac{1}{2}\mathcal{B}\right) \mathrm{d}t + \sum_{j=1}^{N} ia_j\mathcal{B}^{1/2}\mathrm{d}W_t^{(j)}e^{i\phi_t^{(j)}} \\
&= \sum_{j=1}^{N} a_j e^{i\phi_t^{(j)}} \left(-i\omega_0 - \frac{1}{2}\mathcal{B}\right) \mathrm{d}t + V.
\end{aligned}
\tag{3.67}
$$

We now let $N \to +\infty$. Since we have seen that $\lim_{N \to +\infty} V = \mathcal{B}^{1/2}\sigma \mathrm{d}\xi_t$ (upon normalization of the a_j by $N^{1/2}$), we get:

$$\mathrm{d}\gamma_t = (-\omega_0 i - \frac{1}{2}\mathcal{B})\gamma_t \mathrm{d}t + \mathcal{B}^{1/2}\sigma \mathrm{d}\xi_t. \tag{3.68}$$

This is equation (8.60) in [48] (in which $\sigma = \mathcal{B} = 1$). If we reexpress equation (3.68) in real and imaginary parts and make the assumption $\sigma = 1$, we get:

$$\begin{cases} \mathrm{d}\gamma_t^{(R)} = \left(-\frac{1}{2}\mathcal{B}\gamma_t^{(R)} + \omega_0\gamma_t^{(I)}\right)\mathrm{d}t + \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\mathrm{d}W_t^{(R)} \\ \mathrm{d}\gamma_t^{(I)} = \left(-\frac{1}{2}\mathcal{B}\gamma_t^{(I)} - \omega_0\gamma_t^{(R)}\right)\mathrm{d}t + \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\mathrm{d}W_t^{(I)}. \end{cases} \tag{3.69}$$

In the rest of the thesis, we will not use equation (3.69). One of its major limitations is that it assumes that the scatterers are continuously illuminated in time by the source. In practice, this is not true at the light of the way radar systems work (emission of short pulses).

3.2.3 Remarks on the phase model and time scales

Increments in the phase are directly proportional to increments in the position of the scatterer (more precisely the distance). If the increments of the distance between t and $t + \Delta t$ are Gaussian random variables with variance proportional to Δt, we obtain the phase model of the previous section:

$$\mathrm{d}\phi_t^{(n)} = \mathcal{B}^{1/2}\mathrm{d}W_t^{(n)}.$$

Therefore, the phase model used *all over this thesis* implies that the scatterers physically move by Gaussian jumps and independently of one another. This is a much stronger restriction than what the random walk model was assuming, namely uniform phase distribution over $[0, 2\pi[$ at fixed time t.

Theoretically, a brownian motion evolves at all scales (see section 2.1.2), and so the phase itself $\phi_t^{(n)}$ evolves at all time scales. In reality, this is not physically possible, and even if is was, the narrowband approximation which we used in section 1.2.4 to define the reflectivity would then collapse. This must be understood as a limit of modelling the sea clutter (reflectivity) with SDE. The variations for very short timescales should not be taken into account. This sets a limitation for using Field's model with high-frequency time series. If very finely sampled time series of Ψ_t could be recorded (for exemple with $\Delta t = 10^{-5}$ s rather than the 10^{-3} s quoted in chapter 1), it could become irrelevant to use Field's model for very short transitions.

3.3 Field's model as a generalization of the K distribution

x_t, $\gamma_t^{(R)}$ and $\gamma_t^{(I)}$ are stationary processes in Field's theory. They have the following probability densities valid for all t (see chapter 4, see [48] and [122]):

$$\begin{cases} p(x_t = x) = \frac{\alpha^\alpha x^{\alpha-1} e^{-\alpha x}}{\Gamma(\alpha)} \\ p\left(\gamma_t^{(R)} = x\right) = p\left(\gamma_t^{(I)} = x\right) = \frac{1}{\sqrt{\pi}}e^{-x^2}, \end{cases} \tag{3.70}$$

which do not depend on $\mathcal{A}$ nor $\mathcal{B}$.

In this section, we explicit the connection between Field's model [48] and the usual K distribution (see [141]) and observe that Field's model is in agreement with it.

Let

$$\Psi_t^{(C)} = C\Psi_t, \tag{3.71}$$

be the reflectivity of Field's model, Ψ_t, multiplied by a constant C. The reader will soon understand why we do that. The intensity, *i.e.* the squared amplitude of the reflectivity, is:

$$|\Psi_t^{(C)}|^2 = |C\Psi_t|^2 = C^2 x_t(\gamma_t^{(R)\,2} + \gamma_t^{(I)\,2}) \tag{3.72}$$

be the intensity. It can be written:

$$|\Psi_t^{(C)}|^2 = C^2 z_t, \tag{3.73}$$

where $z_t = |\Psi_t|^2$ is the intensity of Field's model. From the first equation of (3.70), we have:

$$p(C^2 x_t = x) = \frac{(\frac{\alpha}{C^2})^\alpha x^{\alpha-1} e^{-\frac{\alpha}{C^2}x}}{\Gamma(\alpha)}. \tag{3.74}$$

This is the gamma distribution for the RCS as expressed in equation (4.24) p 109 of [141], with $b = \frac{\alpha}{C^2}$ and $\nu = \alpha$ where b and ν are usually referred to as the scale and shape parameters respectively. We now understand that with $C = 1$ (Field's model), this gamma distribution would have only one parameter, which is not the general case. It is somewhat more intuitive to work with C^2 rather than b since:

$$\mathbb{E}[C^2 x_t] = C^2. \tag{3.75}$$

From the second equation of (3.70):

$$\sqrt{2}\gamma_t^{(R)} \sim \sqrt{2}\gamma_t^{(I)} \sim \mathcal{N}(0,1), \tag{3.76}$$

and by independence of $\gamma_t^{(R)}$ and $\gamma_t^{(I)}$,

$$(\sqrt{2}\gamma_t^{(R)})^2 + (\sqrt{2}\gamma_t^{(I)})^2 \sim \chi_2^2, \tag{3.77}$$

from which we obtain:

$$\begin{aligned} p\left(\gamma_t^{(R)\,2} + \gamma_t^{(I)\,2} = x\right) &= e^{-x} \\ \Leftrightarrow p\left(u(\gamma_t^{(R)\,2} + \gamma_t^{(I)\,2}) = x\right) &= \frac{e^{-\frac{x}{u}}}{u} \end{aligned} \tag{3.78}$$

for all $u \geq 0$. Then,

$$\begin{aligned} p(C^2 z_t = x) &= \int_0^{+\infty} p(C^2 x_t = u) p(C^2 z_t = x \,|\, C^2 x_t = u) du \\ &= \int_0^{+\infty} \frac{(\frac{\alpha}{C^2})^\alpha x^{\alpha-1} e^{-\frac{\alpha}{C^2}x}}{\Gamma(\alpha)} \frac{e^{-\frac{x}{u}}}{u} du \\ &= \frac{2b^{(\nu+1)/2} x^{(\nu-1)/2}}{\Gamma(\nu)} K_{\nu-1}(2\sqrt{bx}). \end{aligned} \tag{3.79}$$

83

We have retrieved the K distribution for the intensity (see equation (4.26) p 109 of [141]). Field's model is a generalization of the K distribution in the sense that it provides a dynamics extension to it.

The expectation of the non-normalized intensity $C^2 z_t$ is:

$$
\begin{aligned}
\mathbb{E}[C^2 z_t] &= C^2 \mathbb{E}[z_t] = C^2 \mathbb{E}\left[x_t\left(\gamma_t^{(R)\,2} + \gamma_t^{(I)\,2}\right)\right] \\
&= C^2 \mathbb{E}[x_t]\mathbb{E}\left[\gamma_t^{(R)\,2} + \gamma_t^{(I)\,2}\right] \\
&= C^2.
\end{aligned}
\tag{3.80}
$$

We have used independence of x_t and $\gamma_t^{(R)\,2} + \gamma_t^{(I)\,2}$, and the fact that $\mathbb{E}[x_t] = 1$ and $\mathbb{E}[\gamma_t^{(R)\,2} + \gamma_t^{(I)\,2}] = 1$ from equations (3.70) and (3.78). We have our interpretation for C: C^2 is the average intensity of the observed reflectivity. In Field's model, $C = 1$ and we have:

$$
\mathbb{E}[z_t] = 1.
\tag{3.81}
$$

Mathematically speaking, Field's model is normalized because σ^2 defined in equation (3.54) is taken to be one. One way to have this is to choose constant amplitudes equal to 1 for all scatterers, $i.e.$ $a_n = 1$ for all n in equation (3.38). Taking $a_n = C$ for all n yields the above non-normalized reflectivity.

3.4 Summary of Field's model

In this section, we wish to summarize Field's model in one page. Field's model starts from the random walk model for the complex reflectivity of a random medium ($e.g.$ the sea surface). The complex reflectivity, Ψ_t, is also called clutter. It is defined as:

$$
\Psi_t = \lim_{\bar{N} \to +\infty} \sum_{n=1}^{N_t} \frac{a_n}{\bar{N}^{1/2}} e^{i\phi_t^{(n)}},
\tag{3.82}
$$

$i.e.$ the limit of the random walk model when $\bar{N}$ becomes infinite, for a random number of scatterers N_t with expectation $\bar{N}$. To ensure a constant mean power during the limiting procedure, the amplitude of the scatterers, a_n, are normalized by $\bar{N}$. Equation (3.82) can be rewritten:

$$
\Psi_t = \lim_{\bar{N} \to +\infty} \left(\frac{N_t}{\bar{N}}\right)^{1/2} \lim_{\bar{N} \to +\infty} \sum_{n=1}^{N_t} \frac{a_n}{N_t^{1/2}} e^{i\phi_t^{(n)}}.
\tag{3.83}
$$

We denote $x_t = \lim_{\bar{N} \to +\infty} \frac{N_t}{\bar{N}}$. x_t is the RCS (radar cross section) in Field's model, it it positive real-valued. We denote $\gamma_t = \lim_{\bar{N} \to +\infty} \sum_{n=1}^{N_t} \frac{a_n}{N_t^{1/2}} e^{i\phi_t^{(n)}}$. γ_t is the speckle, it is complex-valued. The reflectivity is simply:

$$
\Psi_t = x_t^{1/2} \gamma_t.
\tag{3.84}
$$

The dynamics of the RCS x_t are determined by those of N_t. In Field's model, N_t is a linear Birth-Death-Immigration population model. We can show that x_t is solution to the following $stochastic\ differential\ equation$:

$$
\mathrm{d}x_t = \mathcal{A}(1 - x_t)\mathrm{d}t + \left(2\frac{\mathcal{A}}{\alpha}x_t\right)^{\frac{1}{2}} \mathrm{d}W_t^{(x)},
\tag{3.85}
$$

where $W_t^{(x)}$ is a brownian motion, and $\mathcal{A}$ and α are two constants coming from the population model.

The dynamics of the speckle γ_t are determined by those of the phases $\phi_t^{(n)}$. It is assumed that for each scatterer, the dynamics of the phase is given by:

$$\begin{cases} \mathrm{d}\phi_t^{(n)} = \mathcal{B}^{1/2}\mathrm{d}W_t^{(n)} \\ \phi_0^{(n)} = \Delta^{(n)}, \end{cases} \tag{3.86}$$

where $W_t^{(n)}$ is a brownian motion and $\Delta^{(n)}$ is uniformily distributed over $[0, 2\pi[$. We assume that for $n \neq m$, $\phi_t^{(n)}$ and $\phi_t^{(m)}$ are independent. We can show that under this phase model, if we decompose γ_t in real and imaginary parts as $\gamma_t = \gamma_t^{(R)} + i\gamma_t^{(I)}$, we have:

$$\begin{cases} \mathrm{d}\gamma_t^{(R)} = -\frac{1}{2}\mathcal{B}\gamma_t^{(R)}\mathrm{d}t + \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\mathrm{d}W_t^{(R)} \\ \mathrm{d}\gamma_t^{(I)} = -\frac{1}{2}\mathcal{B}\gamma_t^{(I)}\mathrm{d}t + \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\mathrm{d}W_t^{(I)}, \end{cases} \tag{3.87}$$

where $W_t^{(R)}$ and $W_t^{(I)}$ are independent brownian motions (also independent from $W_t^{(x)}$) and $\mathcal{B}$ is a constant.

Equations (3.84), (3.85) and (3.87) constitute Field's model for the reflectivity. Three constant parameters parameterize the model: $\mathcal{A}$ and α for x_t (from the population model) and $\mathcal{B}$ for γ_t (from the scatterer's phase dynamics). $\mathcal{A}$ and $\mathcal{B}$ have the dimension of a frequency: they are inverse correlation times for x_t and γ_t respectively. α is dimensionless. Field's model is a generalization of the 'static' random walk model (section 1.3 or [141]) when the number of scatterers is negative binomial for infinite number of scatterers. As such, the distribution of $z_t = |\Psi_t|^2$ is a K distribution for all t in Field's model. In addition, it provides the temporal structure of the reflectivity through stochastic differential equations.

Field's model is a bridge between the two preceding chapters since it expresses the sea clutter (chapter 1) in terms of SDE, and as such places it in the framework of stochastic analysis (chapter 2). In the next chapters, we will make developments based on Field's model for the sea clutter, and show that it can provide new tools for the statistical analysis of sea clutter time series and target detection.

Chapter 4

Forward and backward probabilistic inference of the sea clutter

One objective of this thesis is to explore the possibilities of Field's model for synthetic aperture radar imaging (SAR). The unknown behaviour of the sea surface is a problem in SAR imaging because it is based on temporal integration. The synthetic aperture (or antenna) is obtained by the motion of the sensor which illuminates the same point (x, y) of the sea surface (Earth's surface more generally) for about $0.5 - 1$ seconds in the satellite case (see [35]). It is well known that the sea clutter typically decorrelates over shorter durations [140], [43]. The issue is that the integration, which works for static surfaces due to our knowledge of the reflectivity's phase history (assuming constant amplitude), does not work anymore for the sea surface. Indeed, the sea clutter evolves not only because of the space interval but also because of the time interval. Both the phase and amplitude evolve in a pretty unpredictable way. Traditional SAR imaging of the sea surface is therefore problematic, with a blurring effect due to the motion of the surface during the integration interval (see [77], [11], [78]). This is the rationale for what is presented in this chapter. Our idea is that Field's model can give information about the reflectivity's history, which might be usable at some step of the SAR algorithm and could lead to a new concept of a *distribution of SAR images.*

Figure 4.1 represents the integration of several measures by a satellite (oddly similar to Sputnik 1 satellite, the first artificial satellite launched by the soviets in 1957). To every point of view (*i.e.* position) corresponds a sea clutter random process with a stationary distribution (under constant weather conditions). This process can be written $X_t^{(u)}$ where t is the time parameter and u the position of the sensor. Of course, for any u_1, u_2, there is a spatial correlation between $X_t^{(u_1)}$ and $X_t^{(u_2)}$, which starts from 1 if $u_1 = u_2$ and decreases as $|u_1 - u_2|$ increases. This dependency is not treated in this paper. We only make the hypothesis that for all u, the $X_t^{(u)}$ are described by the same model (Field's model, [48], chapter 3). Our objective is to understand the time dependency only, in order to compensate it independently for each of the measures $\tilde{X}_{t_1}^{(u_1)}, \tilde{X}_{t_2}^{(u_2)}, \ldots, \tilde{X}_{t_n}^{(u_n)}$ and bring them together to a common time. Therefore, in this chapter, we will not refer to the space parameter u. We denote simply X_t the sea clutter process from any arbitrarily chosen point of view, and $\tilde{X}_t$ its measures (or realizations). To synchronize the measures, we make probabilistic inferences, based on transition probabilities (see section 4.1).

As explained in section 1.3, statistical models are generally used to describe the sea clutter (Rayleigh distribution, K distribution etc) [141]. These models provide a probability density for the sea clutter X_t that is valid at any fixed time t. They are static in the sense that they do not precisely model the time dependency of the sea clutter and cannot help us

synchronizing the data (unless in a trivial way). On contrary, the SDE representation of the sea clutter presented in chapter 3 (Field's model) is helpful as explained below. We remind that Field's model states that the complex reflectivity is:

$$\Psi_t = x_t^{1/2}(\gamma_t^{(R)} + i\gamma_t^{(I)}), \tag{4.1}$$

with

$$\begin{cases} \mathrm{d}x_t = \mathcal{A}(1 - x_t)\mathrm{d}t + \left(2\frac{\mathcal{A}}{\alpha}x_t\right)^{\frac{1}{2}}\mathrm{d}W_t^{(x)} \\ \mathrm{d}\gamma_t^{(R)} = -\frac{1}{2}\mathcal{B}\gamma_t^{(R)}\mathrm{d}t + \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\mathrm{d}W_t^{(R)} \\ \mathrm{d}\gamma_t^{(I)} = -\frac{1}{2}\mathcal{B}\gamma_t^{(I)}\mathrm{d}t + \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\mathrm{d}W_t^{(I)}. \end{cases} \tag{4.2}$$

In this chapter, we treat systematically several observables: the speckle, the RCS, the intensity and the real reflectivity (real part of the complex reflectivity). The imaginary part of the reflectivity is shown to have the same properties as its real part. One reason is that we cannot say yet what observable is the most important for SAR imaging. One different reason is that SAR imaging is not the only motivation for this chapter. The method could find applications in different settings. For each of these observables, we give mathematical expressions of the transition probabilities from present to future and from present to past which enable respectively forward and backward 'probabilistic inferences'.

In section 4.1, we introduce vocabulary and notations to define what is meant by forward and backward probabilistic inferences, in relation with conditioned probabilities and Markov processes. In section 4.2, we solve the Fokker-Planck equations of the speckle and the RCS, to obtain their transition probabilities, from which we derive also the transition probabilities of the intensity and real part of the reflectivity. We explain how transition probabilities for the full complex reflectivity can be obtained. Numerical simulations are systematically made and give numerical distributions which match the analytical distributions. In section 4.3, we show that reversal of the conditioned probabilities gives identical formula. All previous results, which were valid for forward probabilistic inferences, extend to backward probabilistic inferences. Section 4.4 is a discussion of the applicability of our results to address the problem of carrying measures of the sea clutter from different points of view to the same time, *i.e.* synchronizing. Section 4.5 concludes.

4.1 Forward and backward probabilistic inferences

The speckle and RCS are solutions to SDE given by equation (4.2). A SDE is in the following form:

$$\begin{cases} \mathrm{d}X_t = \mu(X_t, t)\mathrm{d}t + \sigma(X_t, t)\mathrm{d}W_t \\ X_0 = \xi_0 \end{cases} \tag{4.3}$$

where $(W_t)_t$ is a brownian motion, also called Wiener process, and ξ_0 is the initial condition which can be for a time different from 0. μ is called the 'drift' and σ is called the 'volatility'. Under the conditions of Itô's theorem of existence and unicity of the solutions, there exists a unique solution denoted $(X_t)_{t \geq 0}$. We assume that it is real-valued. Let $s, t \in \mathbb{R}^+$ such that $s < t$ and $x \in \mathbb{R}$. The transition probability is the probability measure $A \mapsto p(X_t \in A | X_s = x)$ where A belongs to the Borel σ-algebra, also denoted $p(X_t \in . | X_s = x)$. It should be understood that $p(X_t \in A | X_s = x)$ is an intuitive notation for $\mathbb{E}[\mathbf{1}_A(X_t) | X_s = x]$, which is a conditional expectation (section 2.1.3). If the volatility and drift do not depend explicitely

88
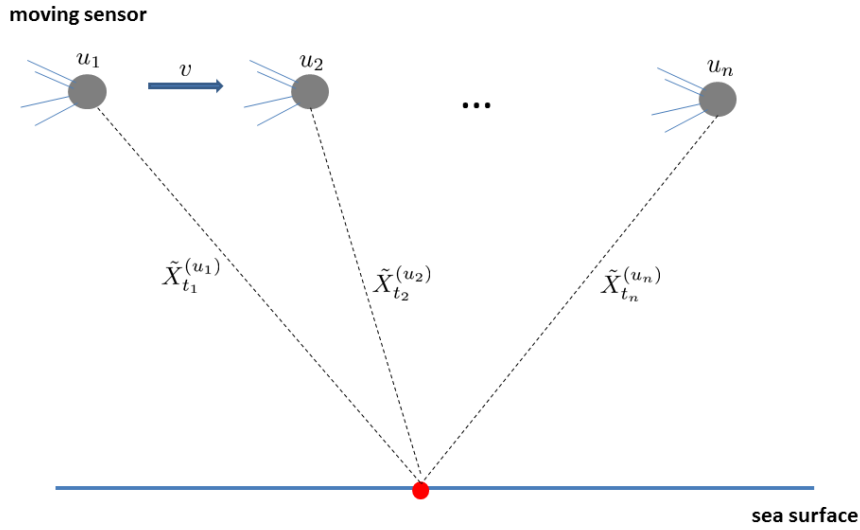
Figure 4.1: Moving sensor measuring the sea clutter $X_t^{(u)}$ at positions $u_1, u_2, \ldots, u_n$ and times $t_1, t_2, \ldots, t_n$. $\tilde{X}_t^{(u)} = X_t^{(u)}(\omega)$ is one realization of the process.

on time, the solution is an homogeneous Markov process (definition 2.12), *i.e.* its transition probabilities depend only on the time interval:

$$p(X_t \in . \,|X_s = x) = p(X_{t-s} \in . \,|X_0 = x). \tag{4.4}$$

The observables we work with: speckle, RCS, intensity and real and imaginary reflectivities, complex reflectivity, are all homogeneous processes. In addition, the speckle, RCS and complex reflectivity are Markov processes. This Markov property is very important. If they were not Markov, the transition probabilities would not simply depend on the observation $X_s = x$ but would depend on the whole past $\{X'_s = x'_s, \ s' \leq s\}$.

The transition probabilities are probabilities of a random variable conditioned by another one, and can be defined from the formalism of conditional expectations. We will refer to them sometimes as transitional and sometimes as conditioned probabilities. It is assumed, for tractability, that for any observable X_t, random vectors $\begin{bmatrix} X_{t_1} & X_{t_2} & \ldots & X_{t_n} \end{bmatrix}^\top$ extracted from the process are absolutely continuous. The same assumption is made if extracted vectors mix up different observables. Defining conditioned probabilities in the most general case is not trivial and care should be taken when dealing with them. However, their calculus rules are quite gentle in the end. In particular, $p(X_t \in . \,|X_s = x)$ is absolutely continuous with a distribution denoted $y \mapsto p(X_t = y|X_s = x)$.

An important result will be used: transformations and conditioning are commutative. More specifically, let X denote a $\mathbb{R}^n$ valued random vector and G a C^1-diffeomorphism between appropriate subsets of $\mathbb{R}^n$. Let Y denote a $\mathbb{R}^m$-valued random vector and $y \in \mathbb{R}^m$. We are again in the framework of absolute continuity. From what was said above, we can consider a random vector denoted $cond(X)$ with the distribution $p(cond(X) = x) = p(X = x \,|\, Y = y)$. Commutativity of conditioning and transformations is expressed by the relation:

$$p(G(X) = x|Y = y) = p(G(cond(X)) = x). \tag{4.5}$$

This result is used in sections 4.2 and 4.3 to compute the conditioned probability of products of independent random variables, which is a special case where we take $G : (x, y) \mapsto (xy, x)$ and then integrate to obtain the conditioned probability of the first component. We think that most physicists would never ask themselves such questions, so if these results are obvious we encourage the reader to skip the corresponding developments.

Besides the mathematical complications, $p(X_t = y|X_s = x)$ can simply be thought of as the 'probability' that X_t is equal to y knowing that X_s is equal to x (with some abuse of terminology). Knowing $p(X_t = y|X_s = x)$ enables what we refer to as a probabilistic inference, *i.e.* a statement of the form "given its value at time s, the sea clutter has a probability p to be in the interval $[a, b]$ at time t". From the deterministic measure $X_s = x$, we can infer the distribution of the possible values for X_t. Since $s < t$, it is a *forward probabilistic inference*. Using Bayes's formula, we can return the conditioned probability to obtain the *backward probabilistic inference* $p(X_s = x|X_t = y)$ (see section 4.3).

4.2 Present to future transition probabilities

Present to future transition probabilities are of the form $p(X_t = x|X_s = y)$ where $s \leq t$. Analytical expressions can be obtained solving the Fokker Planck (a.k.a Kolmogorov forward) equations. In what follows when solving the Fokker Planck equations (FPE), 'p' refers to $p(X_t = x|X_0 = y)$ in the time-dependant case, and to $p(X_\infty = x) = p(X_t = x)$ in the stationary case. We remind that by homogeneity, $p(X_t = x|X_s = y) = p(X_{t-s} = x|X_0 = y)$. In this section, the FPE are expressed and solved for the speckle γ_t and the RCS x_t only. We remind from section 2.3.2 that for a SDE

$$\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t, \tag{4.6}$$

the density function $(x, t) \mapsto p(X_t = x)$ is solution to the following partial differential equation:

$$\frac{\partial p(X_t = x)}{\partial t} = -\frac{\partial}{\partial x}[\mu(x)p(X_t = x)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[\sigma^2(x)p(X_t = x)]. \tag{4.7}$$

The case $p(X_t = x|X_0 = y)$ correspond to the deterministic initial condition $X_0 = y$.

4.2.1 Distributions of the speckle

$\gamma_t^{(R)}$ and $\gamma_t^{(I)}$ are real-valued Ornstein-Uhlenbeck processes (e.g. [105]). They are solution to the same SDE given in equation (4.2), only the driving brownian motion changes. It is therefore sufficient to study $\gamma_t^{(R)}$ for example. From its SDE, we can write its FPE (see section 2.3.2):

$$\frac{\partial p}{\partial t} = \frac{\mathcal{B}}{4}\frac{\partial^2 p}{\partial^2 x} + \frac{1}{2}\mathcal{B}\frac{\partial x\, p}{\partial x}. \tag{4.8}$$

Stationary probability

The stationary FPE for $\gamma_t^{(R)}$ reads:

$$0 = \frac{\mathcal{B}}{4}\frac{\partial^2 p}{\partial^2 x} + \frac{1}{2}\mathcal{B}\frac{\partial x\, p}{\partial x}, \tag{4.9}$$

which can also be written ([120] section 5.2):

$$- \left(-\tfrac{1}{2}\mathcal{B}x\right)p + \frac{\partial}{\partial x}\left(\frac{\mathcal{B}}{4}p\right) = 0$$

$$\Leftrightarrow \frac{-\tfrac{1}{2}\mathcal{B}x}{\mathcal{B}/4}\frac{\mathcal{B}}{4}p = \frac{\partial}{\partial x}\left(\frac{\mathcal{B}}{4}p\right),$$

the solution of which reads [120]:

$$p(x) = \frac{C}{\mathcal{B}/4}\,exp\left(\int_0^x \frac{-\tfrac{1}{2}\mathcal{B}u}{\mathcal{B}/4}du\right)$$

$$\Leftrightarrow p(x) = \frac{C}{\mathcal{B}/4}\,e^{-x^2}.$$

where $C \in \mathbb{R}$ is a constant. Using $\int_{\mathbb{R}} p(x)dx = 1$, this constant is given by $C = \frac{\mathcal{B}}{4\sqrt{\pi}}$. Finally, the stationary distribution is:

$$p\left(\gamma_\infty^{(R)} = x\right) = p\left(\gamma_\infty^{(I)} = x\right) = \frac{1}{\sqrt{\pi}}e^{-x^2}. \tag{4.10}$$

It is a centered Gaussian random variable of variance $1/2$. Since the unconditioned physical speckle is stationary, we have for all $t \geq 0$:

$$p\left(\gamma_t^{(R)} = x\right) = p\left(\gamma_t^{(I)} = x\right) = p\left(\gamma_\infty^{(R)} = x\right). \tag{4.11}$$

Transition probabilities

$\gamma_0^{(R)}$ is the real-part of the speckle at time $t = 0$, whose distribution is given by equation (4.10). Let $y \in \mathbb{R}$ and let assume that we measured $\tilde{\gamma}_0^{(R)} = y$. Taking this present measure into account to predict how the speckle is going to evolve in the future is equivalent to computing the conditioned probability $p\left(\gamma_t^{(R)} = x|\gamma_0^{(R)} = y\right)$, which is solution to the FPE (4.8) with the initial distribution $p = \delta_y$. The Fokker-Planck equation for an Ornstein-Ulhenbeck process is solved p 100 of [120] using the Fourier transform. $\forall x \in \mathbb{R}$, $\forall t > 0$, the solution is given by a Gaussian distribution:

$$p\left(\gamma_t^{(R)} = x|\gamma_0^{(R)} = y\right) = \frac{1}{\sqrt{2\pi v(t)}}e^{-\frac{1}{2}\frac{(x-m_y(t))^2}{v(t)}}, \tag{4.12}$$

with expectation:

$$m_y(t) = ye^{-\mathcal{B}t/2}, \tag{4.13}$$

and variance:

$$v(t) = \frac{1 - e^{-\mathcal{B}t}}{2}. \tag{4.14}$$

The expectation starts from y at $t = 0$ and exponentially decays towards 0 as $t \to +\infty$. The variance starts from 0 at $t = 0$ (Dirac distribution) and increases toward $\frac{1}{2}$ as $t \to +\infty$. Therefore, there is a progressive increase in the uncertainty which nonetheless remains finite as we draw away from the initial condition.

We use the Euler-Maruyama method for solving numerically the SDE for $\gamma_t^{(R)}$ as described in section 2.2.3. Normalized time-dependent histograms are computed from 10000 simulated

trajectories. The results depicted in figure 4.2 show a very accurate agreement between the observed numerical histograms and the analytical distributions given by equation (4.12). We observe the predicted behaviour: exponential decay of the mean of the trajectories toward 0 and progressive increase of their variance toward $\frac{1}{2}$. We can better understand the link between conditioned probabilities and forward probabilistic inferences. For $t = 0.001$ s for example, the distribution is almost centered at the measure $y = 2$ and has a much smaller variance than the asymptotic distribution. We have more constraints on what the measure of $\gamma_{0.001}^{(R)}$ is likely to give than if no measure is taken into account, in which case the asymptotic distribution is the best guess.
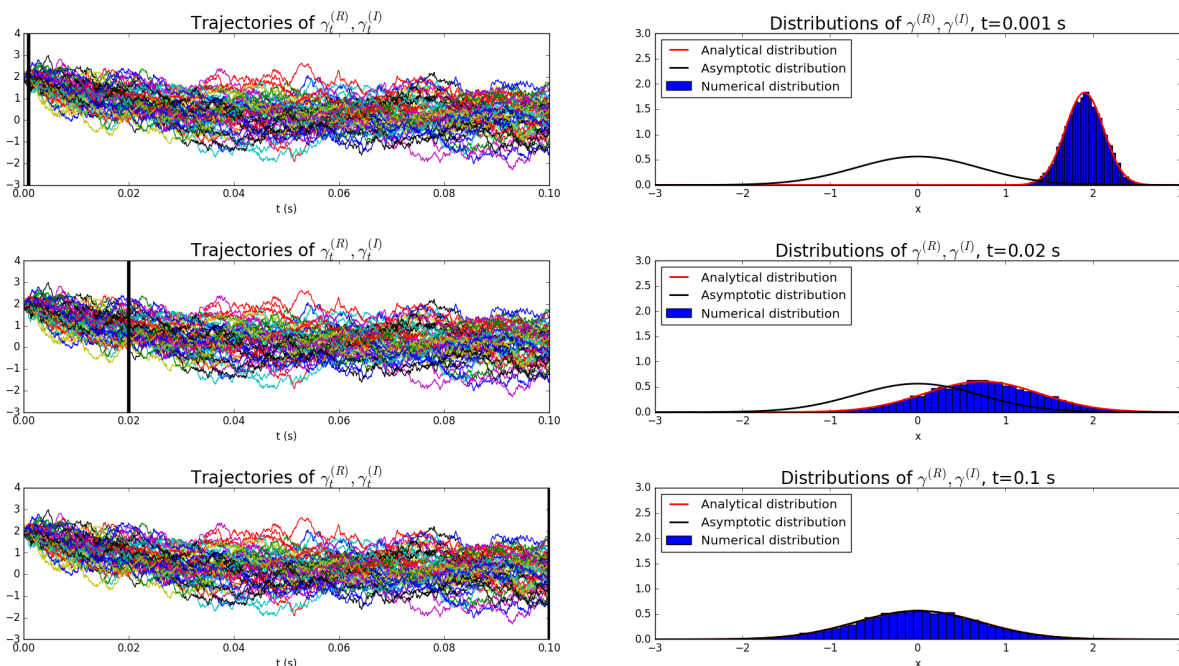
Figure 4.2: Comparison between analytical distributions of $\gamma_t^{(R)}, \gamma_t^{(I)}$ derived from the resolution of the FPE (equation (4.12)), and numerical distributions from the resolutions of the SDE using the Euler-Maruyama method. 10000 trajectories are computed with $\mathcal{B} = 100$ Hz and $y = 2$.

We have performed a Kolmogorov-Smirnov test to quantify the adequation between the numerical and analytical distributions at the three times represented on the right part of figure 4.2: 0.001 s, 0.02 s and 0.1 s. The results are gathered in Table 4.1. The first raw represents the D statistic (maximum distance between the numerical and analytical cumulative density functions) and the second is the p-value. The meaning of the p-value is as follows. Let H_0 be the null hypothesis, which states that the numerical distribution is generated from the analytical one. The p-value is the probability that under H_0, the D statistic is greater or equal than the observed D. For example, at $t = 0.001$ s there is 95.6 % chance that under H_0, the distance between the numerical and analytical distributions is at least 0.016. We cannot reject the null hypothesis and thus the test is positive. The same holds for $t = 0.02$ s and $t = 0.1$ s. A positive Kolmogorov-Smirnov test indicates that our numerical simulations are precise enough, *i.e.* our scheme (Euler-Maruyama) and timestep (0.0001 s here) are good enough.

	$t = 0.001$ s	$t = 0.02$ s	$t = 0.1$ s
D	0.016	0.023	0.016
p-value	0.956	0.649	0.955

Table 4.1: Results of the Kolmogorov-Smirnov test for the transition probabilities of γ_t.

4.2.2 Distributions of the RCS

Following Field in [48], we set $\check{x}_t = \alpha x_t$. This transformation usually results in simpler equations and more tractability. We obtain the following SDE:

$$\mathrm{d}(\check{x}_t) = \mathcal{A}(\alpha - \check{x}_t)\mathrm{d}t + (2\mathcal{A}\check{x}_t)^{\frac{1}{2}}\mathrm{d}W_t^{(x)}. \tag{4.15}$$

The FPE reads:

$$
\begin{aligned}
\frac{\partial p}{\partial t} &= \frac{1}{2}\frac{\partial^2 2\mathcal{A}xp}{\partial^2 x} - \frac{\partial \mathcal{A}(\alpha - x)p}{\partial x} \\
\Leftrightarrow \frac{\partial p}{\partial t} &= \mathcal{A}\frac{\partial}{\partial x}\left(p + x\frac{\partial p}{\partial x}\right) - \mathcal{A}\left(-p + (\alpha - x)\frac{\partial p}{\partial x}\right) \\
\Leftrightarrow \frac{\partial p}{\partial t} &= \mathcal{A}\frac{\partial p}{\partial x} + \mathcal{A}\left(\frac{\partial p}{\partial x} + x\frac{\partial^2 p}{\partial x^2}\right) + \mathcal{A}p - \mathcal{A}(\alpha - x)\frac{\partial p}{\partial x} \\
\Leftrightarrow \frac{\partial p}{\partial t} &= \mathcal{A}x\frac{\partial^2 p}{\partial x^2} + \mathcal{A}(2 - \alpha + x)\frac{\partial p}{\partial x} + \mathcal{A}p.
\end{aligned}
\tag{4.16}
$$

From now on, we assume $\mathcal{A} \neq 0$. If $\mathcal{A} = 0$, the distribution remains identical to the initial condition. If the latter is deterministic (i.e. $p(x,0) = \delta_{x_0}(x)$), we have $\forall\, t \geq 0, p(x,t) = \delta_{x_0}(x)$, which means that the RCS remains constant and equal to x_0.

Stationary probability

The stationary (asymptotic) distribution is solution to:

$$0 = \mathcal{A}x\frac{\partial^2 p}{\partial x^2} + \mathcal{A}(2 - \alpha + x)\frac{\partial p}{\partial x} + \mathcal{A}p. \tag{4.17}$$

It is given p 49 of [48] by:

$$p(\check{x}_\infty = x) = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}\mathbf{1}_{[0,+\infty[}(x), \tag{4.18}$$

where Γ denotes the usual gamma function, and $\mathbf{1}$ the indicator function. Proceeding to the inverse transform $x_t \to \check{x}_t/\alpha$, we get:

$$p(x_\infty = x) = \frac{\alpha(\alpha x)^{\alpha-1}e^{-\alpha x}}{\Gamma(\alpha)}\mathbf{1}_{[0,+\infty[}(x). \tag{4.19}$$

Transition probabilities

x_0 is the RCS at time $t = 0$, whose distribution is given by equation (4.19). Let $y \in \mathbb{R}^+$ and let assume that we measured $\check{x}_0 = y$. Taking this present measure into account to predict how the RCS is going to evolve in the future is equivalent to computing the conditioned probability $p(x_t = x | x_0 = y)$. It is obtained by applying the inverse transform $x_t \to \check{x}_t/\alpha$

to the solution of the FPE (4.16) with a Dirac initial distribution $\delta_{\alpha y}$. A direct resolution of the FPE (4.16) is provided in what follows. We use the asymptotic distribution (4.18) to gain insight into the solution we are looking for, and make the following transformation: $p(x,t) = x^{\alpha-1}\tilde{p}(x,t)$. We get:

$$
\begin{aligned}
x^{\alpha-1}\frac{\partial \tilde{p}}{\partial t} &= \mathcal{A}x\frac{\partial^2 x^{\alpha-1}\tilde{p}}{\partial x^2} + \mathcal{A}(2-\alpha+x)\frac{\partial x^{\alpha-1}\tilde{p}}{\partial x} + \mathcal{A}x^{\alpha-1}\tilde{p} \\
&= \mathcal{A}x\frac{\partial}{\partial x}\left((\alpha-1)x^{\alpha-2}\tilde{p} + x^{\alpha-1}\frac{\partial \tilde{p}}{\partial x}\right) + \mathcal{A}(2-\alpha+x)\left(x^{\alpha-1}\frac{\partial \tilde{p}}{\partial x} + (\alpha-1)x^{\alpha-2}\tilde{p}\right) \\
&\quad + \mathcal{A}x^{\alpha-1}\tilde{p} \\
&= \mathcal{A}x\left((\alpha-1)(\alpha-2)x^{\alpha-3}\tilde{p} + 2(\alpha-1)x^{\alpha-2}\frac{\partial \tilde{p}}{\partial x} + x^{\alpha-1}\frac{\partial^2 \tilde{p}}{\partial x^2}\right) \\
&\quad + \mathcal{A}(2-\alpha+x)\left(x^{\alpha-1}\frac{\partial \tilde{p}}{\partial x} + (\alpha-1)x^{\alpha-2}\tilde{p}\right) + \mathcal{A}x^{\alpha-1}\tilde{p} \\
&= x^{\alpha-1}\mathcal{A}x\frac{\partial^2 \tilde{p}}{\partial x^2} + x^{\alpha-1}\left(2\mathcal{A}(\alpha-1) + \mathcal{A}(2-\alpha+x)\right)\frac{\partial \tilde{p}}{\partial x} \\
&\quad + x^{\alpha-1}\left(\frac{1}{x}\mathcal{A}(\alpha-1)(\alpha-2) + \frac{1}{x}\mathcal{A}(2-\alpha+x)(\alpha-1) + \mathcal{A}\right)\tilde{p} \\
\Leftrightarrow x^{\alpha-1}\frac{\partial \tilde{p}}{\partial t} &= x^{\alpha-1}\mathcal{A}x\frac{\partial^2 \tilde{p}}{\partial x^2} + x^{\alpha-1}\mathcal{A}(\alpha+x)\frac{\partial \tilde{p}}{\partial x} + x^{\alpha-1}\mathcal{A}\alpha\tilde{p} \\
\Leftrightarrow \frac{\partial \tilde{p}}{\partial t} &= \mathcal{A}x\frac{\partial^2 \tilde{p}}{\partial x^2} + \mathcal{A}(\alpha+x)\frac{\partial \tilde{p}}{\partial x} + \mathcal{A}\alpha\tilde{p}. \tag{4.20}
\end{aligned}
$$

To obtain the solution of the FPE, we use the separation of variables $\tilde{p}(x,t) = X(x)T(t)$. The 'prime' symbol refers to derivation with respect to t when it comes after 'T' and with respect to x when it comes after 'X'. We get:

$$
\begin{aligned}
T'(t)X(x) &= \mathcal{A}xT(t)X''(x) + \mathcal{A}(\alpha+x)T(t)X'(x) + \mathcal{A}\alpha T(t)X(x) \tag{4.21} \\
\Leftrightarrow \frac{T'(t)}{T(t)} &= \frac{\mathcal{A}xX''(x) + \mathcal{A}(\alpha+x)X'(x) + \mathcal{A}\alpha X(x)}{X(x)}. \tag{4.22}
\end{aligned}
$$

Thus $\exists \lambda > 0$ such that:

$$
\begin{cases}
T'(t) = -\lambda T(t) \\
\mathcal{A}xX''(x) + \mathcal{A}(\alpha+x)X'(x) + \mathcal{A}\alpha X(x) = -\lambda X(x)
\end{cases}
$$
$$
\Leftrightarrow
$$
$$
\begin{cases}
T(t) = c_\lambda e^{-\lambda t} \\
xX''(x) + (\alpha+x)X'(x) + (\alpha+\frac{\lambda}{\mathcal{A}})X(x) = 0,
\end{cases}
\tag{4.23}
$$

where $c_\lambda \in \mathbb{R}^+$. We get inspiration from the exponential decay of the asymptotic distribution as $x \to +\infty$ and make the transformation $X(x) = z(x)e^{-x}$. We note that:

$$
\begin{cases}
X'(x) = z'(x)e^{-x} - e^{-x}z(x) \\
X''(x) = z''(x)e^{-x} - 2z'(x)e^{-x} + z(x)e^{-x}.
\end{cases}
\tag{4.24}
$$

Injecting these derivatives into the second equation of (4.23) gives:

$$xe^{-x}\left(z''(x) - 2z'(x) + z(x)\right) + (\alpha + x)e^{-x}(z'(x) - z(x)) + \left(\alpha + \frac{\lambda}{\mathcal{A}}\right)e^{-x}z(x) = 0$$

$$\Leftrightarrow \quad xz''(x) + (-2x + (\alpha + x))\,z'(x) + \left(x - (\alpha + x) + \alpha + \frac{\lambda}{\mathcal{A}}\right)z(x) = 0$$

$$\Leftrightarrow \quad xz''(x) + (\alpha - x)z'(x) + \frac{\lambda}{\mathcal{A}}z(x) = 0. \tag{4.25}$$

Equation (4.25) can be written:

$$xz''(x) + (b + 1 - x)z'(x) + az(x) = 0, \tag{4.26}$$

with $b = \alpha - 1$ and $a = \frac{\lambda}{\mathcal{A}}$. Equation (4.26) can equivalently be seen as a Laguerre differential equation or a confluent hypergeometric differential equation. Its two independent solutions are (p 1481 [63]) the generalized Laguerre function $L_a^b(x) = \frac{\Gamma(a+b+1)}{\Gamma(a+1)\Gamma(b+1)}\,_1F_1(-a, b+1, x)$ and the confluent hypergeometric function of the second kind $U(-a, b + 1, x)$ (a.k.a Tricomi's function). For $n_1, n_2 \in \mathbb{N}$, $_{n_1}F_{n_2}$ refers to the hypergeometric function. For more about these functions, refer to [6]. Therefore, the general solution is:

$$X(x) = d_1 L_{\frac{\lambda}{\mathcal{A}}}^{\alpha-1}(x)e^{-x} + d_2 U\left(-\frac{\lambda}{\mathcal{A}}, \alpha, x\right)e^{-x}, \tag{4.27}$$

with $d_1, d_2 \in \mathbb{R}$.

There is an interesting application of the generalized Laguerre functions at section 13.2 of [15]. The resolution of Schrödinger's equation for the hydrogen atom in spherical coordinates using separation of variables gives a Laguerre differential equation for the radial part. It is not explicit in this reference but the confluent hypergeometric function of the second kind is discarded because $U(a, b, 0) = \infty$ if $Re(b) > 1$ [6], which is unacceptable in their context. In our problem, we cannot discard it since $p(x_\infty = 0) = +\infty$ for $\alpha \leq 1$. Again from the asymptotic distribution, we impose an exponential decay for $x \to +\infty$ which leads to a discretization of the possible values of λ.

From [6], $U(-\frac{\lambda}{\mathcal{A}}, \alpha, x) \sim x^{\frac{\lambda}{\mathcal{A}}}$ for $x \to +\infty$, so $d_2 U(-\frac{\lambda}{\mathcal{A}}, \alpha, x)e^{-x} \sim d_2 e^{-x}x^{\frac{\lambda}{\mathcal{A}}}$ for $x \to +\infty$, which means that $\forall \lambda$, it decays exponentially. However, $_1F_1(-\frac{\lambda}{\mathcal{A}}, \alpha, x) \sim e^x x^{-\frac{\lambda}{\mathcal{A}}-\alpha}$ for $x \to +\infty$ if $\frac{\lambda}{\mathcal{A}} \notin \mathbb{N}$, in which case:

$$d_1 L_{\frac{\lambda}{\mathcal{A}}}^{\alpha-1}(x)e^{-x} \sim d_1 \frac{\Gamma(\frac{\lambda}{\mathcal{A}} + \alpha)}{\Gamma(\frac{\lambda}{\mathcal{A}} + 1)\Gamma(\alpha)}x^{-\frac{\lambda}{\mathcal{A}}-\alpha}. \tag{4.28}$$

The decrease is only polynomial for $x \to +\infty$. If $\frac{\lambda}{\mathcal{A}} \in \mathbb{N}$, $_1F_1(-\frac{\lambda}{\mathcal{A}}, \alpha-1, x)$ is a polynomial and $L_{\frac{\lambda}{\mathcal{A}}}^{\alpha-1}$ reduces to the Laguerre polynomial. In that case, $d_1 L_{\frac{\lambda}{\mathcal{A}}}^{\alpha-1}(x)e^{-x}$ decays exponentially for $x \to +\infty$. Under the constraint of exponential decay, we have narrowed down the range of possible λ to:

$$\lambda \in \{0, \mathcal{A}, 2\mathcal{A}, 3\mathcal{A}...\} = \mathcal{A}\mathbb{N}. \tag{4.29}$$

It turns out that for $\lambda \in \mathcal{A}\mathbb{N}$, *i.e.* for $-\frac{\lambda}{\mathcal{A}} \in -\mathbb{N}$, we get [6]:

$$U\left(-\frac{\lambda}{\mathcal{A}}, \alpha, x\right) = U(-n, \alpha, x) = (-1)^n n! L_n^{\alpha-1}(x). \tag{4.30}$$

The general solution of (4.23) reduces to $X(x) = dL_{\frac{\lambda}{\mathcal{A}}}^{\alpha-1}(x)e^{-x}$, where $d \in \mathbb{R}$. The general solution of (4.16) reads:

$$p(x,t) = \sum_{n=0}^{+\infty} c_n e^{-\mathcal{A}nt} e^{-x} x^{\alpha-1} L_n^{\alpha-1}(x). \tag{4.31}$$

It is well known that $\{L_n^{\alpha-1}, n \in \mathbb{N}\}$ is a family of orthogonal polynomial [38]. If $\alpha - 1 > -1$, $\forall n, m \in \mathbb{N}$:

$$\langle L_n^{\alpha-1}, L_m^{\alpha-1} \rangle_{e^{-x} x^{\alpha-1}} = \int_0^{+\infty} L_n^{\alpha-1}(x) L_m^{\alpha-1}(x) e^{-x} x^{\alpha-1} dx = \frac{\Gamma(n+\alpha)}{n!} \delta_{n,m}. \tag{4.32}$$

Using (4.32), we compute the c_n coefficients:

$$
\begin{aligned}
\langle p(.,t), L_k^{\alpha-1} \rangle &= \int_0^{+\infty} \sum_{n=0}^{+\infty} c_n e^{-\mathcal{A}nt} e^{-x} x^{\alpha-1} L_n^{\alpha-1}(x) L_k^{\alpha-1}(x) dx \\
&= \sum_{n=0}^{+\infty} \int_0^{+\infty} c_n e^{-\mathcal{A}nt} e^{-x} x^{\alpha-1} L_n^{\alpha-1}(x) L_k^{\alpha-1}(x) dx \\
&= c_k e^{-\mathcal{A}kt} \frac{\Gamma(k+\alpha)}{k!} \\
\Leftrightarrow c_n &= \langle p(.,t), L_n^{\alpha-1} \rangle e^{\mathcal{A}nt} \frac{n!}{\Gamma(n+\alpha)}
\end{aligned}
$$

At $t = 0$ and with the Dirac initial condition $p(x,0) = \delta_{\alpha y}(x)$, the c_n coefficients reduce to:

$$
\begin{aligned}
c_n &= \langle \delta_{\alpha y}, L_n^{\alpha-1} \rangle \frac{n!}{\Gamma(n+\alpha)} \\
&= L_n^{\alpha-1}(\alpha y) \frac{n!}{\Gamma(n+\alpha)}. \tag{4.33}
\end{aligned}
$$

Replacing the c_n coefficients in (4.31) by their expression (4.33), we obtain that $\forall x \in \mathbb{R}$ and $\forall t > 0$:

$$p(\check{x}_t = x | \check{x}_t = \alpha y) = \sum_{n=0}^{+\infty} \frac{L_n^{\alpha-1}(\alpha y) n!}{\Gamma(n+\alpha)} e^{-\mathcal{A}nt} e^{-x} x^{\alpha-1} L_n^{\alpha-1}(x) \mathbf{1}_{[0,+\infty[}(x). \tag{4.34}$$

This is equation (8.55) of [48], which was given without an explicit proof. Applying the inverse transform $x_t \to \frac{\check{x}_t}{\alpha}$, we finally obtain the transient distributions of the RCS x_t:

$$p(x_t = x | x_0 = y) = \sum_{n=0}^{+\infty} \frac{\alpha L_n^{\alpha-1}(\alpha y) n!}{\Gamma(n+\alpha)} e^{-\mathcal{A}nt} e^{-\alpha x} (\alpha x)^{\alpha-1} L_n^{\alpha-1}(\alpha x). \tag{4.35}$$

We have removed the indicator function because it is always implicit that the distributions of positive random variable, such as x_t, have support in $\mathbb{R}^+$. We use Milstein scheme (definition 2.38) for solving numerically the SDE for x_t as the Euler-Maruyama method revealed itself insufficient: it generated negative values, which is impossible for the RCS, and ended the computation. Normalized time-dependent histograms are computed from 10000 simulated trajectories. In figure 4.3, we observe the progressive variance increase of the conditioned

96
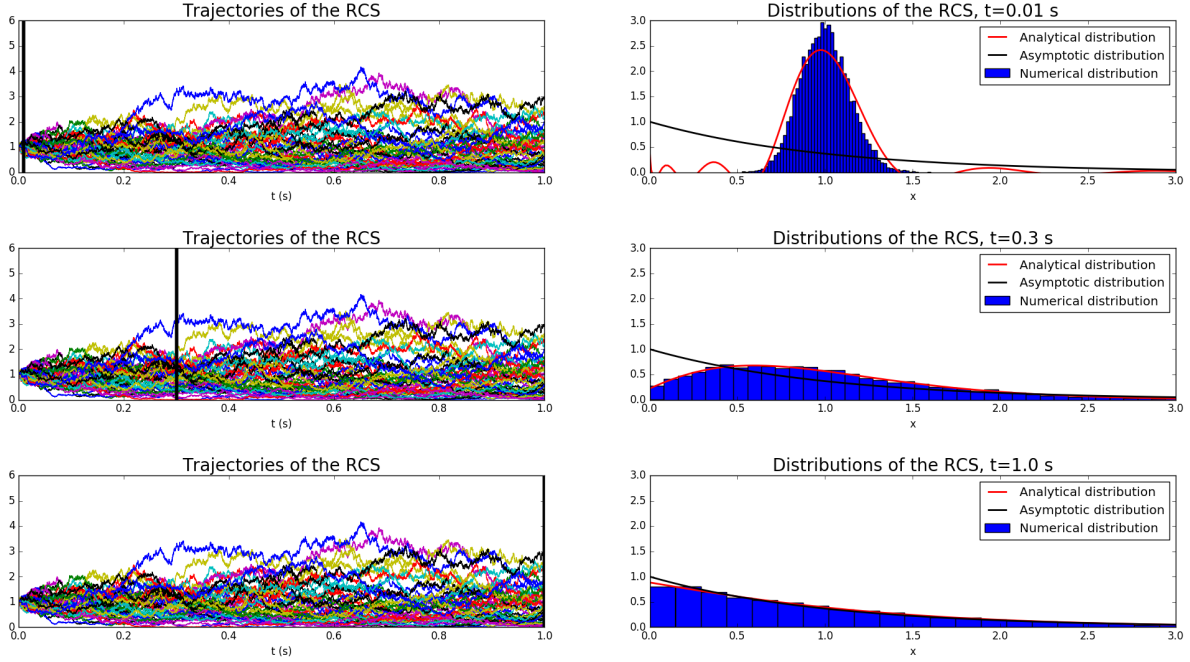
Figure 4.3: Comparison between analytical distributions of x_t derived from the resolution of the FPE, and numerical distributions derived from the resolutions of the SDE using the Milstein method. 10000 trajectories are computed with $\mathcal{A} = 1$ Hz, $\alpha = 1$ and $y = 1$.

distributions, as well as their convergence toward the asymptotic distribution as $t \to +\infty$. It illustrates again the difference between $p(x_t = x|x_0 = y)$ and $p(x_t = x)$, which is asymptotically distributed. There is a gain in using the distribution $p(x_t = x|x_0 = y)$ rather than $p(x_t = x)$ to infer the future measure $\tilde{x}_t$.

In figure 4.3, one can also note that there is an accurate agreement between the numerical and analytical distributions. However, oscillations appears in the analytical solution for t close to 0, which can be due to the fact that we were able to compute the sum in (4.35) up to $n = 150$ only. That is the approximate limit of Python for computing $n!$ and $\Gamma(n+\alpha)$. To go beyond, one could maybe approximate the ratio $\frac{n!}{\Gamma(n+\alpha)}$ which should evolve much more slowly than $n!$ and $\Gamma(n+\alpha)$ since $n! = \Gamma(n+1)$. Another issue is the computing time for evaluating the Laguerre polynomials for large n. Though we have proved a formula for the transition probability of x_t, we have somewhat reinvented the wheel! Indeed, x_t is a Cox-Ingersoll-Ross process. It has been shown that its transition probabilities can be written analytically [46]:

$$p(x_t = x|x_0 = y) = ce^{-cx-c\delta y} \left(\frac{x}{y\delta}\right)^{\frac{\alpha-1}{2}} I_{\alpha-1}(2c\sqrt{xy\delta}), \qquad (4.36)$$

with $c = \frac{\alpha}{1-e^{-\mathcal{A}t}}$ and $\delta = e^{-\mathcal{A}t}$, and where $I_{\alpha-1}$ is the modified Bessel function of the first kind of order $\alpha - 1$. This result is merely cited in [48] p 63, we will not try to prove the equivalence of equations (4.35) and (4.36). It could be interesting to look into [143] to find elements to prove the equivalence. We did numerical simulations in the same way as those represented in figure 4.3 with the difference that this time, equation (4.36) is used for the exact transition probabilities. The results are represented in figure 4.4. The oscillations

	Laguerre Polynomials			Bessel function		
	$t = 0.001$ s	$t = 0.02$ s	$t = 0.1$ s	$t = 0.001$ s	$t = 0.02$ s	$t = 0.1$ s
D	0.0358	0.019	0.019	0.019	0.037	0.025
p-value	0.149	0.875	0.872	0.848	0.127	0.558

Table 4.2: Results of the Kolmogorov-Smirnov test for the transition probabilities of x_t.

observed in figure 4.3 have disappeared and the fit between the numerical and analytical distributions is even better.
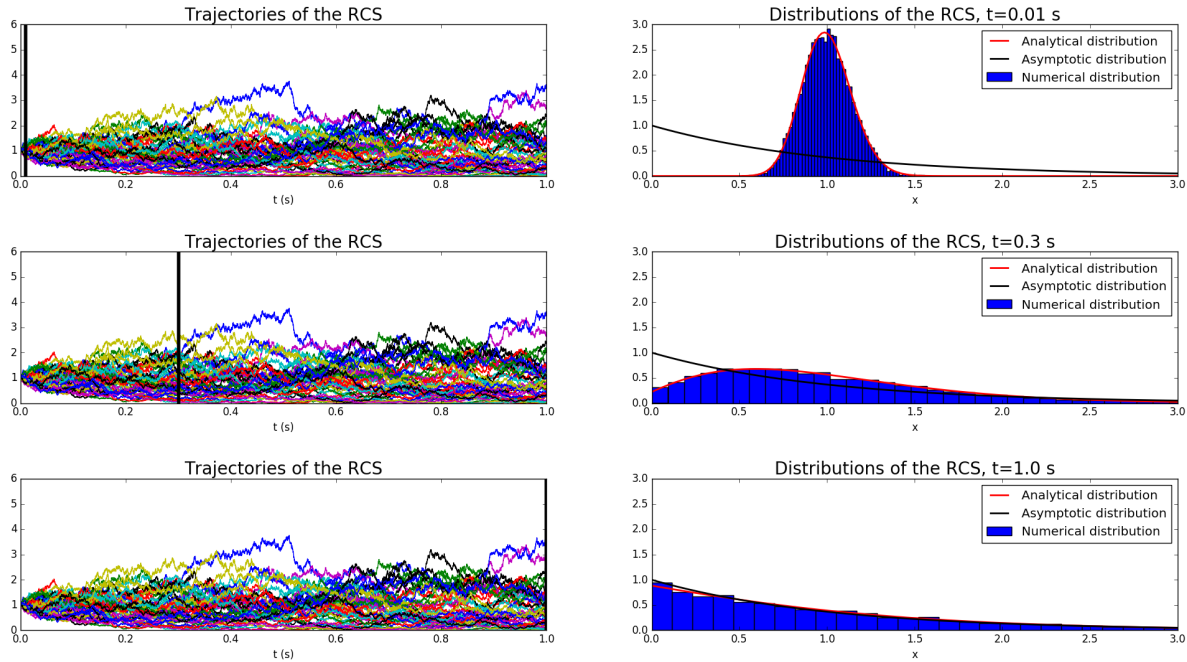
Figure 4.4: Comparison between analytical distributions of x_t expressed with the Bessel function (equation (4.36)), and numerical distributions derived from the resolutions of the SDE using the Milstein method. 10000 trajectories are computed with $\mathcal{A} = 1$ Hz, $\alpha = 1$ and $y = 1$.

As for the speckle, we have performed a Kolmogorov-Smirnov test to quantify the adequation between the numerical and analytical distributions at the three times represented on the right part of figures 4.3 and 4.4: 0.01 s, 0.3 s and 1 s. The results are gathered in Table 4.2. The interpretation of the results is the same as for the speckle. One difference is that we tested against two different analytical distributions, the one expressed in terms of Laguerre polynomials given by equation (4.35) (though truncated at $n = 150$), and the one expressed with the Bessel function given by equation (4.36). Some p-values are relatively low, but they stay above the significance level of 5 %. The positivity of the Kolmogorov-Smirnov test indicates again that our numerical simulations are precise enough, *i.e.* our scheme (Milstein) and timestep (0.0001 s here) are good enough.

4.2.3 Distributions of the real (and imaginary) reflectivity

The real and imaginary parts of the reflectivity are simply defined by $R_t = Re(\Psi_t) = x_t^{1/2}\gamma_t^{(R)}$ and $I_t = Im(\Psi_t) = x_t^{1/2}\gamma_t^{(I)}$. Since $\gamma_t^{(R)}$ and $\gamma_t^{(I)}$ have the same distributions (for equal initial conditions), so will R_t and I_t. Thus, it is sufficient to compute those of R_t. We remind that the processes $x_t^{1/2}$ and $\gamma_t^{(R)}$ are independent. In the following, we denote g the function $g : x \mapsto x^{1/2}$.

Stationary probability

From $x_\infty^{1/2} = g(x_\infty)$ and from equation (4.19) we get:

$$
\begin{aligned}
p(x_\infty^{1/2} = x) &= p(x_\infty = x^2)2x \\
&= \frac{2\alpha^\alpha \, x^{2\alpha-1} \, e^{-\alpha x^2}}{\Gamma(\alpha)}.
\end{aligned}
\tag{4.37}
$$

We compute $p(R_\infty = x)$ as the distribution of the product of 2 independent random variables:

$$
\begin{aligned}
p(R_\infty = x) &= p\left(x_\infty^{1/2}\gamma_\infty^{(R)} = x\right) \\
&= \int_{\mathbb{R}+} p(x_\infty^{1/2} = u)p\left(\gamma_\infty^{(R)} = x/u\right)\frac{1}{u}du \\
&= \frac{2\alpha^\alpha}{\sqrt{\pi}\Gamma(\alpha)} \int_{\mathbb{R}+} u^{2\alpha-2}e^{-\left(\left(\frac{x}{u}\right)^2+\alpha u^2\right)}du.
\end{aligned}
\tag{4.38}
$$

Transition probabilities

$R_t = x_t^{1/2}\gamma_t^{(R)}$ is the product of the square-root of the RCS and the speckle. We will show that we can compute $p\left(R_t = x|x_0 = y, \gamma_0^{(R)} = z\right)$ from the transition probabilities $p(\gamma_t^{(R)} = . |\gamma_0^{(R)} = .)$ and $p(x_t = . |x_0 = .)$. From $x_t^{1/2} = g(x_t)$ and from equation (4.35), we can compute the conditioned probability $p(x_t^{1/2} = x|x_0 = y)$. Indeed, we remind that transformations and conditioning are commutative. As a result, the conditioned probability $p(x_t^{1/2} = x|x_0 = y)$ is the probability of the transformation by the function g of a random variable of conditioned probability $p(x_t = x|x_0 = y)$:

$$
\begin{aligned}
p(x_t^{1/2} = x|x_0 = y) &= p(g(x_t) = x|x_0 = y) \\
&= p(x_t = x^2|x_0 = y)2x \\
\Leftrightarrow \\
p(x_t^{1/2} = x|x_0 = y) &= \sum_{n=0}^{+\infty} \frac{2\alpha L_n^{\alpha-1}(\alpha y)n!}{\Gamma(n+\alpha)}e^{-\mathcal{A}nt}xe^{-\alpha x^2}(\alpha x^2)^{\alpha-1}L_n^{\alpha-1}(\alpha x^2).
\end{aligned}
\tag{4.39}
$$

By independence of the processes x_t and $\gamma_t^{(R)}$ and from the properties of conditioned

probabilities in the framework of absolute continuity:

$$
\begin{aligned}
p\left(x_t^{1/2} = x | x_0 = y,\ \gamma_0^{(R)} = z\right) &= \frac{p\left(x_t^{1/2} = x,\ x_0 = y,\ \gamma_0^{(R)} = z\right)}{p\left(x_0 = y,\ \gamma_0^{(R)} = z\right)} \\
&= \frac{p\left(x_t^{1/2} = x,\ x_0 = y\right) p\left(\gamma_0^{(R)} = z\right)}{p(x_0 = y)p\left(\gamma_0^{(R)} = z\right)} \\
&= \frac{p\left(x_t^{1/2} = x,\ x_0 = y\right)}{p(x_0 = y)} \\
&= p\left(x_t^{1/2} = x | x_0 = y\right),
\end{aligned}
\tag{4.40}
$$

which is very intuitive. Similarly, $p\left(\gamma_t^{(R)} = x | x_0 = y,\ \gamma_0^{(R)} = z\right) = p\left(\gamma_t^{(R)} = x | \gamma_0^{(R)} = z\right)$. Again by commutativity of transformations and conditioning, we can compute

$$
p\left(x_t^{1/2}\gamma_t^{(R)} = x | x_0 = y,\ \gamma_0^{(R)} = z\right)
\tag{4.41}
$$

as the distribution of the product of 2 independent random variables:

$$
p\left(x_t^{1/2}\gamma_t^{(R)} = x | x_0 = y,\ \gamma_0^{(R)} = z\right) = \int_0^{+\infty} p\left(\gamma_t^{(R)} = x/u \mid x_0 = y,\ \gamma_0^{(R)} = z\right)
$$
$$
p\left(x_t^{1/2} = u \mid x_0 = y,\ \gamma_0^{(R)} = z\right) \frac{1}{u}du
$$

$\Leftrightarrow$

$$
p\left(R_t = x \mid x_0 = y,\ \gamma_0^{(R)} = z\right) = \int_0^{+\infty} \frac{1}{\sqrt{2\pi v(t)}} e^{-\frac{1}{2}\frac{(\frac{x}{u} - m_z(t))^2}{v(t)}} \sum_{n=0}^{+\infty} \frac{2\alpha L_n^{\alpha-1}(\alpha y)n!}{\Gamma(n+\alpha)}
\tag{4.42}
$$
$$
e^{-Ant}e^{-\alpha u^2}(\alpha u^2)^{\alpha-1}L_n^{\alpha-1}(\alpha u^2)du
$$

where $m_z(t)$, $v(t)$ are expressed in equations (4.13) and (4.14). The mathematical details are provided in appendix B and are based on relation (4.5) applied to the couple $(x_t^{1/2}, \gamma_t^{(R)})$.

Equation (4.42) is an exact analytical expression of the transition probabilities of R_t. It is explicit and relatively easy to evaluate numerically. As mentioned earlier, it is also valid for I_t where the condition $\gamma_0^{(R)} = z$ is replaced by $\gamma_0^{(I)} = z$. Figure 4.5 shows the very accurate agreement between the analytical and numerical distributions. We solve numerically the SDE for $x_t, \gamma_t^{(R)}$ with Dirac initial conditions and then compute R_t with the relation $R_t = x_t^{1/2}\gamma_t^{(R)}$. As for x_t, oscillations appear for short times since the sum (4.35) must be truncated approximately at $n = 150$. We observe a difference between the transition probabilities of x_t and $\gamma_t^{(R)}$ on one side, and those of R_t on the other side. x_t was conditioned only by x_0 and $\gamma_t^{(R)}$ by $\gamma_0^{(R)}$, but R_t is conditioned by x_0 and $\gamma_0^{(R)}$, not just R_0. It is more constraining since $R_0 = x_0^{1/2}\gamma_0^{(R)}$. We observed numerically that the transition probabilities cannot depend only on R_0 but must depend on both x_0 and $\gamma_0^{(R)}$. For example, the conditions $x_0 = 1$, $\gamma_0^{(R)} = 2$ and $x_0 = 4$, $\gamma_0^{(R)} = 1$ give different transitional probabilities even though in both cases $R_0 = 2$. Physically, this is explained by the difference between the dynamics of the speckle and the RCS, which evolve on different timescales. The same remark holds in the next section for the intensity.
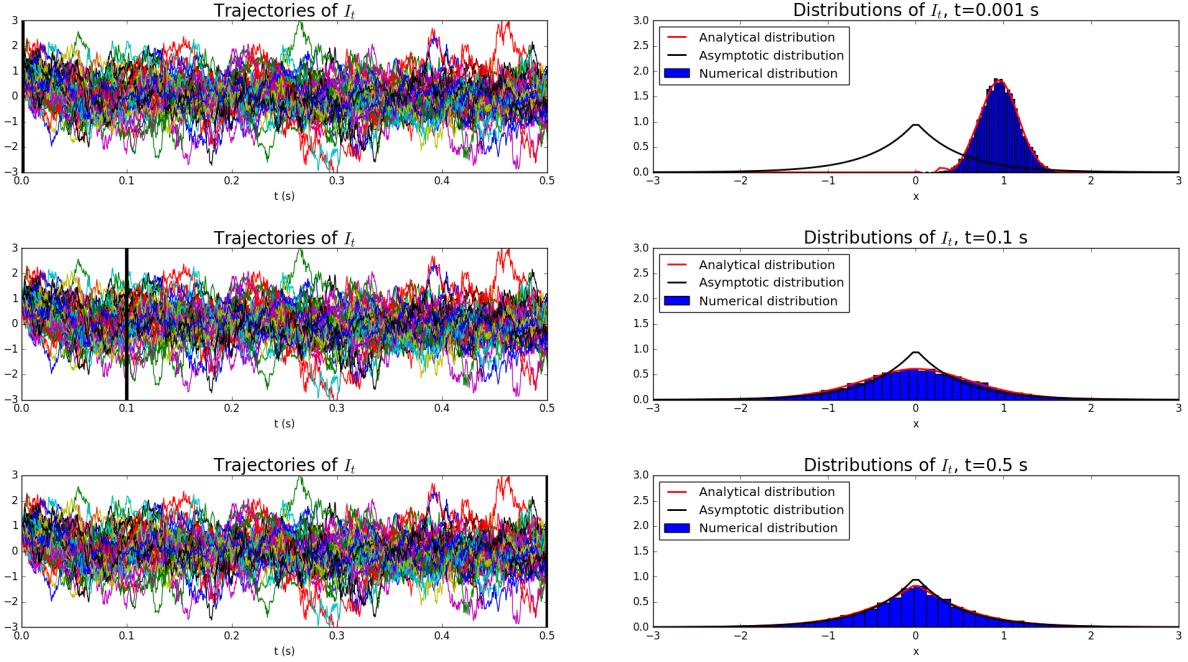
Figure 4.5: Comparison between analytical distributions of R_t or I_t (4.42) and numerical distributions. 10000 trajectories are computed with $\mathcal{A} = 1$ Hz, $\alpha = 1$, $\mathcal{B} = 100$ Hz, $x_0 = 1$ and $\gamma_0^{(R)} = 1$.

4.2.4 Distributions of the intensity

The intensity z_t is defined by $z_t = |\Psi_t|^2 = x_t \left(\gamma_t^{(R)\,2} + \gamma_t^{(I)\,2} \right)$. In this section, we derive its transition probabilities.

Stationary probability

The stationary distribution of the intensity is the classical K distribution as shown below. Equation (4.10) states that $\gamma_\infty^{(R)} \sim \gamma_\infty^{(I)} \sim \mathcal{N}(0, 1/2)$, i.e. $\sqrt{2}\gamma_\infty^{(R)} \sim \sqrt{2}\gamma_\infty^{(I)} \sim \mathcal{N}(0, 1)$. Since $\gamma_\infty^{(R)}$ and $\gamma_\infty^{(I)}$ are independent, $2\gamma_\infty^{(R)\,2} + 2\gamma_\infty^{(I)\,2} \sim \chi_2^2$. We get:

$$
\begin{aligned}
p\left(2\gamma_\infty^{(R)\,2} + 2\gamma_\infty^{(I)\,2} = x\right) &= \frac{1}{2}e^{-x/2} \\
\Leftrightarrow p\left(\gamma_\infty^{(R)\,2} + \gamma_\infty^{(I)\,2} = x\right) &= e^{-x}
\end{aligned}
\tag{4.43}
$$

x_∞ and $\gamma_\infty^{(R)\,2} + \gamma_\infty^{(I)\,2}$ are independent and their distributions are given respectively by equation (4.19) and equation (4.43). As the distribution of a product of 2 independent random variables, we get:

$$
\begin{aligned}
p(z_\infty = x) &= p\left(x_\infty \left(\gamma_\infty^{(R)\,2} + \gamma_\infty^{(I)\,2}\right) = x\right) \\
&= \int_{\mathbb{R}^+} p(x_\infty = u)\, p\left(\left(\gamma_\infty^{(R)\,2} + \gamma_\infty^{(I)\,2}\right) = \frac{x}{u}\right)\frac{1}{u}du \\
&= \int_{\mathbb{R}^+} \frac{\alpha(\alpha u)^{\alpha-1}e^{-\alpha u}}{\Gamma(\alpha)}\frac{e^{-x/u}}{u}du.
\end{aligned}
$$

Since $\frac{e^{-x/u}}{u} = p(z_\infty = x | x_\infty = u)$ ([141] p 103) and $\frac{\alpha(\alpha u)^{\alpha-1} e^{-\alpha u}}{\Gamma(\alpha)} = p(x_\infty = u)$ we get (see [141] p 109):

$$p(z_\infty = x) = \frac{2\alpha^{\frac{\alpha+1}{2}} x^{\frac{\alpha-1}{2}}}{\Gamma(\alpha)} K_{\alpha-1}(2\sqrt{\alpha x}), \tag{4.44}$$

where K is the modified Bessel function of the second kind.

Transition probabilities

We know from section 4.2.1 that $\forall x \in \mathbb{R}$, $\forall t > 0$:

$$\begin{cases} p\left(\gamma_t^{(R)} = x | \gamma_0^{(R)} = z\right) = \frac{1}{\sqrt{2\pi v(t)}} e^{-\frac{1}{2} \frac{(x - m_z(t))^2}{v(t)}} \\ p\left(\gamma_t^{(I)} = x | \gamma_0^{(I)} = w\right) = \frac{1}{\sqrt{2\pi v(t)}} e^{-\frac{1}{2} \frac{(x - m_w(t))^2}{v(t)}} \end{cases} \tag{4.45}$$

with:

$$\begin{cases} v(t) = \frac{1 - e^{-\mathcal{B}t}}{2} \\ m_z(t) = z e^{-\frac{\mathcal{B}}{2}t} \\ m_w(t) = w e^{-\frac{\mathcal{B}}{2}t}. \end{cases} \tag{4.46}$$

Let $X = \frac{1}{\sqrt{v(t)}} \begin{pmatrix} cond\left(\gamma_t^{(R)}\right) \\ cond\left(\gamma_t^{(I)}\right) \end{pmatrix}$ where $cond\left(\gamma_t^{(R)}\right)$ and $cond\left(\gamma_t^{(I)}\right)$ are independent random variables such that $p\left(cond\left(\gamma_t^{(R)}\right) = x\right) = p\left(\gamma_t^{(R)} = x | \gamma_0^{(R)} = z\right)$ and $p\left(cond\left(\gamma_t^{(I)}\right) = x\right) = p\left(\gamma_t^{(I)} = x | \gamma_0^{(I)} = z\right)$. Then $\mathbb{E}[X] = \frac{1}{\sqrt{v(t)}} \begin{pmatrix} m_R(t) \\ m_I(t) \end{pmatrix}$ and from the independence of $cond\left(\gamma_t^{(R)}\right)$ and $cond\left(\gamma_t^{(I)}\right)$, $\Gamma_X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ where Γ_X is the covariance matrix of X. We apply theorem 1.3.4 p 22 of [102], with $n = 2$ and a non-centrality coefficient δ:

$$\begin{aligned} \delta(t) &= \frac{1}{\sqrt{v(t)}} \begin{pmatrix} m_R(t) & m_I(t) \end{pmatrix} \frac{1}{\sqrt{v(t)}} \begin{pmatrix} m_R(t) \\ m_I(t) \end{pmatrix} \\ &= \frac{1}{v(t)} (m_R(t)^2 + m_I(t)^2) \\ &= \frac{2e^{-\mathcal{B}t}(z^2 + w^2)}{1 - e^{-\mathcal{B}t}}. \end{aligned}$$

We obtain:

$$p\left(\frac{1}{v(t)}\left(cond\left(\gamma_t^{(R)}\right)^2 + cond\left(\gamma_t^{(I)}\right)^2\right) = x\right) = \frac{1}{2} e^{-\frac{x + \delta(t)}{2}} {}_0F_1\left(1, \frac{1}{4}\delta(t)x\right).$$

We can express this result with the modified Bessel function of the first kind $\mathcal{I}_0$ defined as:

$$\mathcal{I}_0(z) = \sum_{n=0}^{+\infty} \frac{\left(\frac{1}{4}z^2\right)^n}{n! \Gamma(n+1)}. \tag{4.47}$$

Upon replacing $z = \sqrt{\delta(t)x}$ we get:

$$I_0(\sqrt{\delta(t)x}) = \sum_{n=0}^{+\infty} \frac{\left(\frac{1}{4}\delta(t)x\right)^n}{n!^2} = {}_0F_1\left(1, \frac{1}{4}\delta(t)x\right)$$

such that

$$p\left(\frac{1}{v(t)}\left(cond\left(\gamma_t^{(R)}\right)^2 + cond\left(\gamma_t^{(I)}\right)^2\right) = x\right) = \frac{1}{2}e^{-\frac{x+\delta(t)}{2}}I_0\left(\sqrt{\delta(t)x}\right).$$

Let g be the function:

$$g: \quad \mathbb{R} \quad \to \mathbb{R}$$
$$x \quad \mapsto v(t)x$$

We have $cond\left(\gamma_t^{(R)}\right)^2 + cond\left(\gamma_t^{(I)}\right)^2 = g\left(\frac{1}{v(t)}\left(cond\left(\gamma_t^{(R)}\right)^2 + cond\left(\gamma_t^{(I)}\right)^2\right)\right)$ and we obtain:

$$p\left(cond\left(\gamma_t\right)^2 = x\right) = \frac{1}{2v(t)}e^{-\frac{\frac{x}{v(t)}+\delta(t)}{2}}I_0\left(\sqrt{\frac{\delta(t)x}{v(t)}}\right),$$

with $cond\left(\gamma_t\right)^2 = cond\left(\gamma_t^{(R)}\right)^2 + cond\left(\gamma_t^{(I)}\right)^2$. We also introduce the notation $\gamma_t^2 = \gamma_t^{(R)\,2} + \gamma_t^{(I)\,2}$.

We define $cond(x_t)$ as a random variable such that $p(cond(x_t) = x) = p(x_t = x|x_0 = y)$, for example solution to the RCS for x_t in equation (4.2) with a Dirac-distributed initial condition. $cond(x_t)$ and $cond\left(\gamma_t\right)^2$ are independent, from what we get:

$$p\left(cond\left(x_t\right)cond\left(\gamma_t\right)^2 = x\right) = \int_0^{+\infty} p\left(cond\left(\gamma_t\right)^2 = x/u\right)p\left(x_t = u\right)\frac{1}{u}du$$

$\Leftrightarrow$

$$p\left(z_t = x|x_0 = y, \gamma_0^{(R)} = z, \gamma_0^{(I)} = w\right) = \int_0^{+\infty} \frac{1}{2v(t)}e^{-\frac{\frac{x}{v(t)u}+\delta(t)}{2}}I_0\left(\sqrt{\frac{\delta(t)x}{v(t)u}}\right)$$
$$\sum_{n=0}^{+\infty} \frac{\alpha L_n^{\alpha-1}(\alpha y)n!}{\Gamma(n+\alpha)}e^{-\mathcal{A}nt}e^{-\alpha u}(\alpha u)^{\alpha-1}L_n^{\alpha-1}(\alpha u)\frac{1}{u}du. \tag{4.48}$$

We obtained equation (4.48) by application of transformations to conditioned random variables, which is equivalent to conditioning the transformed random variables, according to relation (4.5). We will not provide the full details since we want to maintain the focus on the results. The proof is similar to that of the real (and imaginary) reflectivity in appendix B. Equation (4.48) yields the distribution of z_t conditioned by the measure $x_0 = y$, $\gamma_0^{(R)} = z$, $\gamma_0^{(I)} = w$. However, the knowledge of both $\gamma_t^{(R)}$ and $\gamma_t^{(I)}$ is not necessary. z and w take part in the expression of $\delta(t)$ only, in which only the value $z^2 + w^2$ must be known. Consequently, one can state that

$$p\left(z_t = x|x_0 = y, \gamma_0^{(R)\,2} + \gamma_0^{(I)\,2} = u\right) = \int_0^{+\infty} \frac{1}{2v(t)}e^{-\frac{\frac{x}{v(t)u}+\delta(t)}{2}}I_0\left(\sqrt{\frac{\delta(t)x}{v(t)u}}\right)$$
$$\sum_{n=0}^{+\infty} \frac{\alpha L_n^{\alpha-1}(\alpha y)n!}{\Gamma(n+\alpha)}e^{-\mathcal{A}nt}e^{-\alpha u}(\alpha u)^{\alpha-1}L_n^{\alpha-1}(\alpha u)\frac{1}{u}du, \tag{4.49}$$

where $u = z^2 + w^2$. Equation (4.49) is an exact analytical expression of the distribution of z_t which is explicit and relatively easy to evaluate numerically. Assuming that the RCS is constant, we get $x_t = 1$ or $p(x_t = x) = \delta_1(x)$. Replacing this expression in (4.49) gives after calculations:

$$
\begin{aligned}
p(z_t = x | x_t = s) &= \frac{1}{(1 - e^{-\mathcal{B}t})s} e^{-\frac{x + z_0 e^{-\mathcal{B}t}}{(1 - e^{-\mathcal{B}t})s}} I_0 \left(\sqrt{\frac{4 e^{-\mathcal{B}t} z_0 x}{(1 - e^{-\mathcal{B}t})^2 s^2}} \right) \\
&= \frac{1}{(1 - e^{-\mathcal{B}t})s} e^{-\frac{x + z_0 e^{-\mathcal{B}t}}{(1 - e^{-\mathcal{B}t})s}} I_0 \left(\frac{2 e^{-\mathcal{B}\frac{t}{2}}}{(1 - e^{-\mathcal{B}t})s} \sqrt{x z_0} \right),
\end{aligned}
$$

which is formula 8.53 p 63 of [48]. Equation (4.49) is therefore a generalization of this formula for RCS varying in time according to Field's model.

Figure 4.6 represents numerical trajectories of the intensity. We simulated trajectories of x_t using the Miltein method and trajectories of $\gamma_t^{(R)}$ and $\gamma_t^{(I)}$ using the Euler-Maruyama method. The intensity was then computed using the relation $z_t = x_t \left(\gamma_t^{(R)\,2} + \gamma_t^{(I)\,2} \right) = x_t \gamma_t^2$. The initial conditions were: $x_0 = 1$, $\gamma_0^{(R)} = 1$, $\gamma_0^{(I)} = 1$, i.e. $z_0 = 2$. There is a very good agreement between formula (4.49) and numerical distributions. The distribution is almost centered around $x = 2$ at $t = 0.001$ s and then progressively converges toward the asymptotic K distribution (formula (4.44)). As for x_t, I_t, Q_t, oscillations appear for short times since the sum in equation (4.35) must be truncated approximately at $n = 150$.
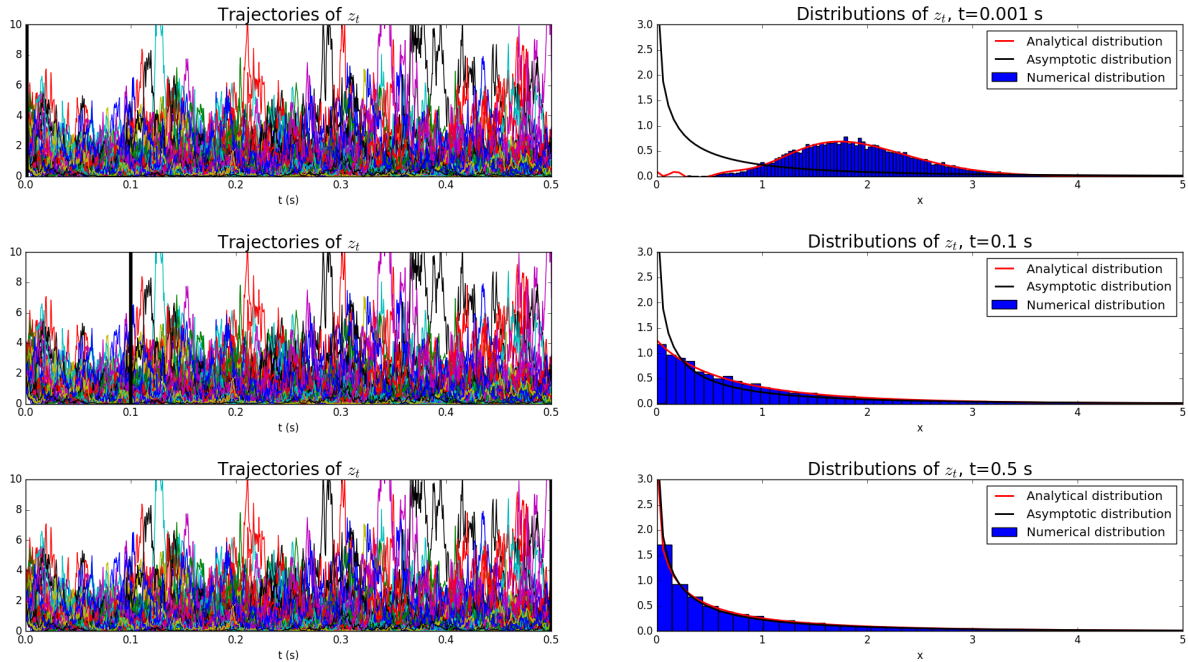

Figure 4.6: Comparison between analytical distributions of z_t (4.49) and numerical distributions. 10000 trajectories are computed with $\mathcal{A} = 1$ Hz, $\alpha = 1$, $\mathcal{B} = 100$ Hz, $x_0 = 1$, $\gamma_0^{(R)} = 1$, $\gamma_0^{(I)} = 1$.

4.2.5 Distributions of the complex reflectivity

We can express the transition probabilities of the complex reflectivity Ψ_t, or more precisely of a non-ambiguous representation of it. If we choose the RCS and speckle coordinate system (the most simple one):

$$\Psi_t \cong \begin{bmatrix} x_t & \gamma_t^{(R)} & \gamma_t^{(I)} \end{bmatrix}^\top. \tag{4.50}$$

We emphasize that three coordinates are necessary to fully 'understand' the reflectivity. Indeed, we could think that $\begin{bmatrix} R_t & I_t \end{bmatrix}^\top$ is a good representation of Ψ_t since $\Psi_t = R_t + iI_t$. However, R_t and I_t are not diffusion processes: they do not solve SDE in the form $\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t$. Yet, a SDE can be written for them but the drift and volatility will not depend on R_t (resp. I_t) uniquely, but on two processes, for example x_t and $\gamma_t^{(R)}$ (resp. x_t and $\gamma_t^{(I)}$) since $R_t = x_t^{1/2}\gamma_t^{(R)}$. This fact can also be compared to what we noticed for the transition probabilities of z_t and R_t (or I_t): there is an ambiguity if we condition one of these by only one observed value. To get rid of the ambiguity, we require the observation of x_t, as seen in equations (4.42) and (4.49). Therefore, three diffusion processes are required to describe the dynamics of Ψ_t which explains equation (4.50). There is some latitude regarding the choice of the coordinate system, since for example:

$$\Psi_t \cong \begin{bmatrix} x_t & R_t & I_t \end{bmatrix}^\top \tag{4.51}$$

is a different possible representation of Ψ_t.

By independence of x_t, $\gamma_t^{(R)}$, $\gamma_t^{(I)}$, we now feel free to write directly (see the few lines which led to equation (4.40)):

$$p\left(\left(x_t, \gamma_t^{(R)}, \gamma_t^{(I)}\right) = (x, y, z) \mid \left(x_0, \gamma_0^{(R)}, \gamma_0^{(I)}\right) = (x', y', z')\right)$$
$$= p\left(x_t = x \mid x_0 = x'\right) p\left(\gamma_t^{(R)} = y \mid \gamma_0^{(R)} = y'\right) p\left(\gamma_t^{(I)} = z \mid \gamma_0^{(I)} = z'\right). \tag{4.52}$$

From equations (4.12) and (4.36), we get:

$$p\left(\left(x_t, \gamma_t^{(R)}, \gamma_t^{(I)}\right) = (x, y, z) \mid \left(x_0, \gamma_0^{(R)}, \gamma_0^{(I)}\right) = (x', y', z')\right)$$
$$= \frac{1}{2\pi v(t)} e^{-\frac{1}{2v(t)}\left[(y - m_{y'}(t))^2 - (z - m_{z'}(t))^2\right]} c e^{-cx - c\delta x'} \left(\frac{x}{x'\delta}\right)^{\frac{\alpha-1}{2}} I_{\alpha-1}(2c\sqrt{xx'\delta}). \tag{4.53}$$

We wish to compare numerical distributions and the analytical distribution (4.53), and to represent them in a way similar to what was done previously (figure 4.2 to 4.6). For graphical representability purposes, we set $x_t = 1$, i.e. we make the hypothesis of a constant RCS. In that case, the initial condition $x_0 = x'$ must be $x_0 = 1$ and the transition probability for x_t is a Dirac distribution:

$$p\left(x_t = x \mid x_0 = x'\right) = \delta_1. \tag{4.54}$$

Ψ_t is now represented simply by $\begin{bmatrix} \gamma_t^{(R)} & \gamma_t^{(I)} \end{bmatrix}^\top$, and equation (4.53) becomes:

$$p\left(\left(\gamma_t^{(R)}, \gamma_t^{(I)}\right) = (y, z) \mid \left(\gamma_0^{(R)}, \gamma_0^{(I)}\right) = (y', z')\right) = \frac{1}{2\pi v(t)} e^{-\frac{1}{2v(t)}\left[(y - m_{y'}(t))^2 - (z - m_{z'}(t))^2\right]},$$

which is simply a 2D Gaussian distribution. To illustrate the progressive spreading (increasing variance) and the concordance between numerical simulations and the analytical

distribution, we simulated 1000 trajectories of $\begin{bmatrix} \gamma_t^{(R)} & \gamma_t^{(I)} \end{bmatrix}^\top$ with $B = 100$ Hz, all starting from $\begin{bmatrix} 0 & 0 \end{bmatrix}^\top$. The results, in figure 4.7, represent the numerical (left) and analytical (right) distributions at time 0.001 s, 0.005 s and 0.02 s. We observe the progressive spreading of the Gaussian distributions, until it reaches approximately its asymptotic distribution at $t = 0.02$ s. It is now manifest why the solution of a SDE is called a diffusion process: the initial dirac is indeed diffusing away, as in a thermal or a chemical diffusion, or even the erosion of a mountain!

Figure 4.8 represents 100 trajectories of the vector process $\begin{bmatrix} \gamma_t^{(R)} & \gamma_t^{(I)} \end{bmatrix}^\top$. We can see the starting point $[0\ 0]^\top$ to the left where all the trajectories initiate. The trajectories are represented for t between 0 s and 0.02 s, and the vertical semi-transparent planes are at $t = 0.001$ s, $t = 0.005$ s and $t = 0.02$ s. The distributions of the intersection points of the trajectories with the vertical planes correspond to the histograms represented in figure 4.7.

4.3   Present to past transition probabilities

At the stage we have reached, we are able to perform forward probabilistic predictions (present to future) for $x_t, \gamma_t^{(R)}, \gamma_t^{(I)}, R_t, I_t, z_t$ and the complex reflectivity, or vector process $\begin{bmatrix} x_t & \gamma_t^{(R)} & \gamma_t^{(I)} \end{bmatrix}^\top$. Let X_t denote any of these and assume that a measure $\tilde{X}_t = x$ is made at time t. In section 4.2, we derived forward conditioned probabilities of the type $p(X_t = x|X_0 = y)$ for the RCS and speckle, and $p(X_t = x|Y_0 = y, Z_0 = z)$ for the real (and imaginary) reflectivity and the intensity. To answer the question "what was the value of X_{t-h} knowing that $X_t = x$ or that $Y_t = y$, $Z_t = z$?", we must reverse the conditions to obtain $p(X_{t-h} = y|X_t = y)$ or $p(X_{t-h} = x|Y_t = y, Z_t = z)$, the distribution of X_{t-h} conditioned by measures located in the future relative to $t - h$. The resulting distributions can be used to make backward probabilistic inferences (present to past). In the remaining of this section, we treat the RCS and speckle together, then the real (and imaginary) reflectivity, the intensity, and the complex reflectivity in different sections.

4.3.1   Distributions of the speckle and RCS

Reversing the present to future probabilities for the speckle, equation (4.12), and for the RCS, equation (4.35), is straightforward as shown below. Let X_t denote either the speckle or the RCS. We can write the transition probabilities in the following way:

$$
\begin{aligned}
p(X_{t-h} = y|X_t = x) &= \frac{p(X_t = x,\ X_{t-h} = y)}{p(X_t = x)} \\
&= \frac{p(X_t = x|X_{t-h} = y)p(X_{t-h} = y)}{p(X_t = x)} \\
&= \frac{p(X_t = x|X_{t-h} = y)p(X_\infty = y)}{p(X_\infty = x)},
\end{aligned} \tag{4.55}
$$

where we have used Bayes formula. We have also used absolute continuity and the fact that the physical process is asymptotically distributed at any time t.

In the case of the speckle, the asymptotic distribution is given by formula (4.10) from which:

$$
\frac{p(\gamma_\infty^{(R)} = y)}{p(\gamma_\infty^{(R)} = x)} = e^{-(y^2 - x^2)}. \tag{4.56}
$$

106
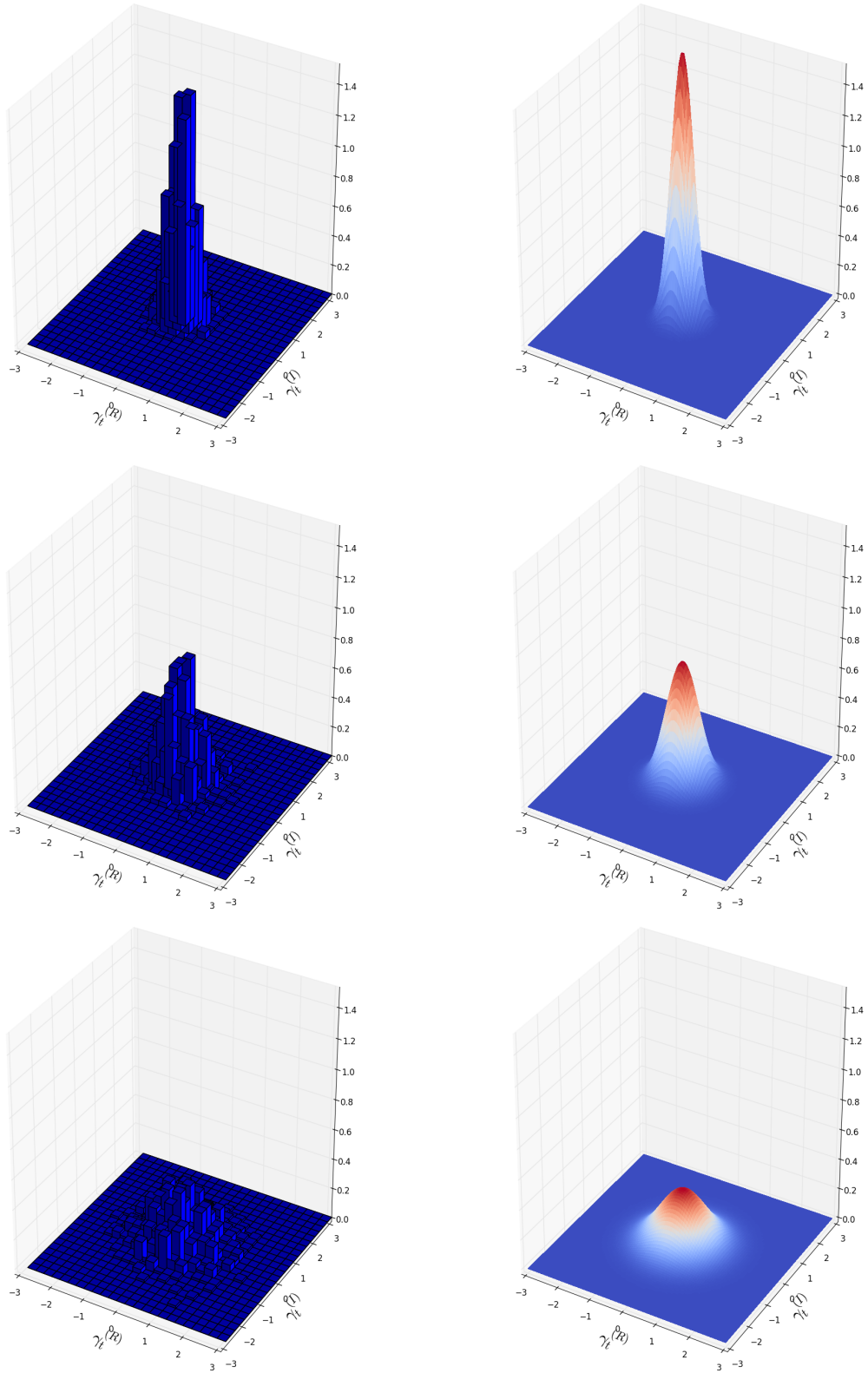
Figure 4.7: Comparison between analytical distributions of $[\gamma_t^{(R)}\ \gamma_t^{(I)}]^\top$ and numerical distributions. 1000 trajectories are computed with $\mathcal{B} = 100$ Hz, $\gamma_0^{(R)} = 0$, $\gamma_0^{(I)} = 0$. *Left*: numerical distributions. *Right*: analytical distributions. Top to bottom: $t = 0.001,\ 0.005,\ 0.02$ s.
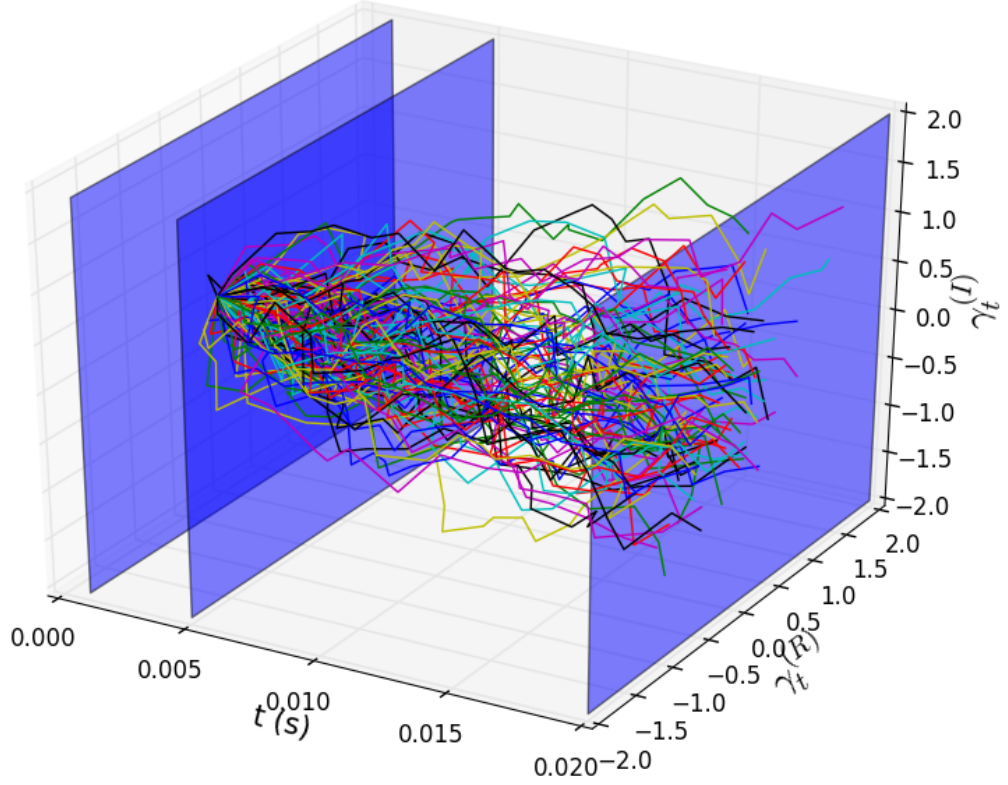
Figure 4.8: Representation of 100 trajectories of $[\gamma_t^{(R)} \ \gamma_t^{(I)}]^\top$ from $t = 0$ s to $t = 0.02$ s. $\mathcal{B} = 100$ Hz and $\gamma_0^{(R)} = 0$, $\gamma_0^{(I)} = 0$.

We also remind that by homogeneity of the Markov process $\gamma_t^{(R)}$, $p(\gamma_t^{(R)} = x | \gamma_{t-h}^{(R)} = y) = p(\gamma_h^{(R)} = x | \gamma_0^{(R)} = y) = \frac{1}{\sqrt{2\pi v(h)}} e^{-\frac{1}{2} \frac{(x - m(h))^2}{v(h)}}$ with $v(h) = \frac{1 - e^{-\mathcal{B}h}}{2}$ and $m(h) = y e^{-\mathcal{B}h/2}$. After some calculations, we can show that:

$$
\begin{aligned}
p(\gamma_{t-h}^{(R)} = y | \gamma_t^{(R)} = x) &= \frac{1}{\sqrt{2\pi v(h)}} e^{-\frac{1}{2} \frac{(x - y e^{-\mathcal{B}h/2})^2}{v(h)}} e^{-(y^2 - x^2)} \\
&= \frac{1}{\sqrt{2\pi v(h)}} e^{-\frac{1}{2} \frac{(y - x e^{-\mathcal{B}h/2})^2}{v(h)}} \\
\Leftrightarrow p(\gamma_{t-h}^{(R)} = y | \gamma_t^{(R)} = x) &= p(\gamma_h^{(R)} = y | \gamma_0^{(R)} = x).
\end{aligned}
\tag{4.57}
$$

In the case of the RCS, the asymptotic distribution is given by formula (4.19) from which:

$$
\frac{p(x_\infty = y)}{p(x_\infty = x)} = \left(\frac{y}{x} \right)^{\alpha - 1} e^{-\alpha(y - x)}.
\tag{4.58}
$$

Using the forward transition probabilities (4.35) and the homogeneity of the process x_t, we

108

obtain:

$$p(x_{t-h} = y | x_t = x) = \sum_{n=0}^{+\infty} \frac{\alpha L_n^{\alpha-1}(\alpha y) n!}{\Gamma(n+\alpha)} e^{-\mathcal{A}nt} e^{-\alpha x} \left(\frac{\alpha xy}{x}\right)^{\alpha-1} L_n^{\alpha-1}(\alpha x) e^{-\alpha(y-x)}$$

$$= \sum_{n=0}^{+\infty} \frac{\alpha L_n^{\alpha-1}(\alpha x) n!}{\Gamma(n+\alpha)} e^{-\mathcal{A}nt} e^{-\alpha y} (\alpha y)^{\alpha-1} L_n^{\alpha-1}(\alpha y)$$

$$\Leftrightarrow p(x_{t-h} = y | x_t = x) = p(x_h = y | x_0 = x) \tag{4.59}$$

Formula (4.57) and (4.59) show that for the speckle and the RCS, the same formula hold for backward and forward probabilistic inferences.

4.3.2 Distributions of the real (and imaginary) reflectivity

In section 4.2.3, we have obtained the distribution of R_h conditioned by $(x_0 = y, \gamma_0^{(R)} = z)$, which is the same as the distribution of R_t conditioned by $(x_{t-h} = y, \gamma_{t-h}^{(R)} = z)$. We would like to obtain the distribution of R_{t-h} conditioned by $(x_t = y, \gamma_t^{(R)} = z)$. To do so, we first reverse the conditioning of the couple $\left(x_t, \gamma_t^{(R)}\right)$:

$$p\left(x_t = z, \gamma_t^{(R)} = w | x_{t-h} = x, \gamma_{t-h}^{(R)} = y\right) = \frac{p\left(x_t = z, x_{t-h} = x, \gamma_t^{(R)} = w, \gamma_{t-h}^{(R)} = y\right)}{p\left(x_{t-h} = x, \gamma_{t-h}^{(R)} = y\right)}$$

$$= \frac{p\left(x_t = z, x_{t-h} = x\right) p\left(\gamma_t^{(R)} = w, \gamma_{t-h}^{(R)} = y\right)}{p(x_{t-h} = x) p\left(\gamma_{t-h}^{(R)} = y\right)}$$

$$= p(x_t = z | x_{t-h} = x) p\left(\gamma_t^{(R)} = w | \gamma_{t-h}^{(R)} = y\right).$$

We have used the independence of the processes x_t and $\gamma_t^{(R)}$ in the second equality. Similarly, we can show that:

$$p\left(x_{t-h} = z, \gamma_{t-h}^{(R)} = w | x_t = x, \gamma_t^{(R)} = y\right) = p(x_{t-h} = z | x_t = x) p\left(\gamma_{t-h}^{(R)} = w | \gamma_t^{(R)} = y\right).$$

Since

$$p(x_{t-h} = z | x_t = x) = p(x_h = z | x_0 = x) = p(x_t = z | x_{t-h} = x) \tag{4.60}$$

and

$$p\left(\gamma_{t-h}^{(R)} = w | \gamma_t^{(R)} = y\right) = p\left(\gamma_t^{(R)} = w | \gamma_{t-h}^{(R)} = y\right), \tag{4.61}$$

we get:

$$p\left(x_{t-h} = z, \gamma_{t-h}^{(R)} = w | x_t = x, \gamma_t^{(R)} = y\right) = p\left(x_t = z, \gamma_t^{(R)} = w | x_{t-h} = x, \gamma_{t-h}^{(R)} = y\right). \tag{4.62}$$

The relation (4.62) is similar to relations (4.57) and (4.59) for the couple $\left(x_t, \gamma_t^{(R)}\right)$ seen as a $\mathbb{R}^2-$valued process. By commutativity of C^1-diffeomorphisms and conditioning, $p\left(R_{t-h} = z | x_t = x, \gamma_t^{(R)} = y\right)$ is obtained as the distribution of the appropriate transformation (see section 4.2.3) applied to the couple $\left(x_{t-h}, \gamma_{t-h}^{(R)}\right)$ conditioned by $\left(x_t = x, \gamma_t^{(R)} = y\right)$,

which has the conditioned distribution $p\left(x_{t-h} = z, \gamma_{t-h}^{(R)} = w | x_t = x, \gamma_t^{(R)} = y\right)$. From formula (4.62), we obtain exactly the same result as we did in section 4.2.3:

$$p\left(R_{t-h} = x | x_t = y, \gamma_t^{(R)} = z\right) = p\left(R_h = x | x_0 = y, \gamma_0^{(R)} = z\right) \qquad (4.63)$$

Formula (4.63) along with formula (4.42) enables backward probabilistic inferences of the real (and imaginary) reflectivity.

4.3.3 Distributions of the intensity

In section 4.2.4, we have obtained the distribution of z_h conditioned by

$$(x_0 = y, \ \gamma_0^{(R)\,2} + \gamma_0^{(I)\,2} = u), \qquad (4.64)$$

which by homogeneity is the same as the distribution of z_t conditioned by

$$\left(x_{t-h} = y, \ \gamma_{t-h}^{(R)\,2} + \gamma_{t-h}^{(I)\,2} = u\right). \qquad (4.65)$$

We would like to obtain the distribution of z_{t-h} conditioned by $(x_t = y, \ \gamma_t^{(R)\,2} + \gamma_t^{(I)\,2} = u)$. To do so, we reverse the conditioning of the random vector $\left(x_t, \gamma_t^{(R)}, \gamma_t^{(I)}\right)$. By mutual independence of x_t, $\gamma_t^{(R)}$ and $\gamma_t^{(I)}$, we can show that:

$$
\begin{aligned}
p \quad & \left(x_t = x, \gamma_t^{(R)} = y, \gamma_t^{(I)} = z | x_{t-h} = u, \gamma_{t-h}^{(R)} = v, \gamma_{t-h}^{(I)} = w\right) \\
= \quad & p(x_t = x | x_{t-h} = u) p\left(\gamma_t^{(R)} = y | \gamma_{t-h}^{(R)} = v\right) p\left(\gamma_t^{(I)} = z | \gamma_{t-h}^{(I)} = w\right) \\
= \quad & p(x_{t-h} = x | x_t = u) p\left(\gamma_{t-h}^{(R)} = y | \gamma_t^{(R)} = v\right) p\left(\gamma_{t-h}^{(I)} = z | \gamma_t^{(I)} = w\right) \\
= \quad & p\left(x_{t-h} = x, \gamma_{t-h}^{(R)} = y, \gamma_{t-h}^{(I)} = z | x_t = u, \gamma_t^{(R)} = v, \gamma_t^{(I)} = w\right).
\end{aligned}
\qquad (4.66)
$$

Formula (4.49) resulted from transformations that implicitely led from

$$p\left(x_t = .\,, \gamma_t^{(R)} = .\,, \gamma_t^{(I)} = . \ | x_{t-h} = y, \gamma_{t-h}^{(R)} = z, \gamma_{t-h}^{(I)} = w\right) \qquad (4.67)$$

to

$$p\left(x_t \left(\gamma_t^{(R)\,2} + \gamma_t^{(I)\,2}\right) = . \ | x_{t-h} = y, \gamma_{t-h}^{(R)\,2} + \gamma_{t-h}^{(I)\,2} = u\right). \qquad (4.68)$$

It is not explicit since the transformations were applied to the marginal distributions

$$p\left(x_t = . \ | x_{t-h} = y, \gamma_{t-h}^{(R)} = z, \gamma_{t-h}^{(I)} = w\right), \qquad (4.69)$$

$$p\left(\gamma_t^{(R)} = . \ | x_{t-h} = y, \gamma_{t-h}^{(R)} = z, \gamma_{t-h}^{(I)} = w\right), \qquad (4.70)$$

etc. This approach was justified by the commutativity between conditioning and transformations.

Similarly, the distribution

$$p\left(x_{t-h} \left(\gamma_{t-h}^{(R)\,2} + \gamma_{t-h}^{(I)\,2}\right) = . \ | x_t = y, \gamma_t^{(R)\,2} + \gamma_t^{(I)\,2} = u\right) \qquad (4.71)$$

will be obtained from the same transformations applied to

$$p\left(x_{t-h} = . , \gamma_{t-h}^{(R)} = . , \gamma_{t-h}^{(I)} = . \,| x_t = u, \gamma_t^{(R)} = v, \gamma_t^{(I)} = w\right). \tag{4.72}$$

Using equation (4.66) and the homogeneity of the intensity, we get a relation similar to formula (4.57), (4.59) and (4.63) for the intensity:

$$p\left(z_{t-h} = x | x_t = y, \gamma_t^{(R)\,2} + \gamma_t^{(I)\,2} = u\right) = p\left(z_h = x | x_0 = y, \gamma_0^{(R)\,2} + \gamma_0^{(I)\,2} = u\right). \tag{4.73}$$

Formula (4.73) along with formula (4.49) enables backward probabilistic inferences of the intensity.

4.3.4   Distributions of the complex reflectivity

As for R_t and z_t, we can reverse the transition probabilities of the vector process $\begin{bmatrix} x_t & \gamma_t^{(R)} & \gamma_t^{(I)} \end{bmatrix}^\top$ (representation of Ψ_t) directly:

$$p\left(\left(x_{t-h}, \gamma_{t-h}^{(R)}, \gamma_{t-h}^{(I)}\right) = (x, y, z) \mid \left(x_t, \gamma_t^{(R)}, \gamma_t^{(I)}\right) = (x', y', z')\right)$$
$$= \quad p\left(\left(x_h, \gamma_h^{(R)}, \gamma_h^{(I)}\right) = (x, y, z) \mid \left(x_0, \gamma_0^{(R)}, \gamma_0^{(I)}\right) = (x', y', z')\right). \tag{4.74}$$

Formula (4.73) along with formula (4.49) enables backward probabilistic inferences of the complex reflectivity in the coordinate system $\begin{bmatrix} x_t & \gamma_t^{(R)} & \gamma_t^{(I)} \end{bmatrix}^\top$.

4.4   Discussion

4.4.1   Observability of x_t

From sections 4.2 and 4.3 we know how to perform forward and backward probabilistic inferences for $\gamma_t^{(R)}, \gamma_t^{(I)}, x_t, R_t, I_t, z_t$. We have left aside a few comments that we address here. Formula (4.12), (4.35), (4.42), (4.49) and (4.53) are for forward probabilistic inferences. We remind that the processes at stake here are all homogeneous processes: the formula apply to any forward leap no matter the starting time. Formula (4.57), (4.59), (4.63), (4.73) and (4.74) are for backward probabilistic inferences. They are directly expressed for any starting time t, for a backward leap of length h.

It was assumed that we can actually measure the starting values in practice. Measuring it gives a condition $X_t = y$ that we can project forward to time X_{t+h} or backward to time X_{t-h}, with $h > 0$. The assumption is justified directly for some quantities: the radar which observes the sea surface records a time series of the complex-valued reflectivity Ψ_t. Taking the real and imaginary parts respectively gives R_t and I_t, and taking the squared-modulus gives z_t. The phase θ_t can also directly be obtained by taking an argument of Ψ_t. However, the RCS x_t is not directly observed. Fayard and Field provide formula to optimally estimate it from increments of z_t and of the phase θ_t [44]. Once the time series of x_t has been obtained from this algorithm, $\gamma_t^{(R)}$ and $\gamma_t^{(I)}$ can be computed from $\gamma_t^{(R)} = \frac{R_t}{x_t^{1/2}}$ and $\gamma_t^{(I)} = \frac{I_t}{x_t^{1/2}}$. The issue is that first it is quite hard to understand what they do, since no explicit formula is given for the estimator of x_t, and second the estimation requires to know the parameters

$\mathcal{A}, \alpha$. The estimation of $\mathcal{A}$ and α is treated in the next chapter. However, the reader will realize that it is necessary to observe x_t in the first place to estimate these parameters. A solution to this circular estimation problem is proposed in chapter 7.

4.4.2 Synchronizing measures

Figure 4.9 represents schematically a discrete time series of values of the RCS (for example) measured by a moving sensor (or different sensors) from positions u_1, u_2, u_3, u_4, u_5 at times t_1, t_2, t_3, t_4, t_5. To be compared, these measures must be transported to the same common time, chosen to be $t = 0.25$ s here, $i.e.$ the time at which the central measure $\tilde{X}_{t_3}^{(u_3)}$ has been taken. The measures $\tilde{X}_{t_1}^{(u_1)}$ and $\tilde{X}_{t_2}^{(u_2)}$ are projected forward using formula (4.35). The projection of $\tilde{X}_{t_1}^{(u_1)}$ at time $t_1 + h$ can be seen as a random variable with the distribution $p\left(X_{t_1+h}^{(u_1)} = . \,|X_{t_1}^{(u_1)} = \tilde{X}_{t_1}^{(u_1)}\right)$, which has an increasing variance with h as seen for example in figure 4.3. This is represented in figure 4.9 by the two solid blue lines diverging from $\tilde{X}_{t_1}^{(u_1)}$. The variance increases until it reaches a maximum value of $1/\alpha$ for $h = +\infty$ (see formula (4.19) of the asymptotic distribution). The expectation converges toward 1 as $h \rightarrow +\infty$. Those are asymptotic values which are not reached if h is not too large. We saw in figure 4.3 that for $A = 1$ Hz and $\alpha = 1$, the asymptotic distribution is not reached for $h = 0.25$ s but it is reached for $h = 0.5$ s. In the example of figure 4.9, the projection of the deterministic measure $\tilde{X}_{t_1}^{(u_1)}$ at the time reference $t = 0.25$ s is a random variable with the distribution $p\left(X_{t_1+0.25}^{(u_1)} = . \,|X_{t_1}^{(u_1)} = \tilde{X}_{t_1}^{(u_1)}\right)$, whose variance is smaller than the asymptotic variance $1/\alpha = 1$ and expectation different from the asymptotic expectation 1. This distribution is our best guess of what the measure of the RCS from position u_1 at the reference time t_3 would be. In the same way, projecting forward $\tilde{X}_{t_2}^{(u_2)}$ to the time $t_3 = 0.25$ s would give a distribution which would be our best guess of what the measure of the RCS from position u_2 would be at the reference time t_3. The above explanations apply as well for backward projection. Backward projection of the measures $\tilde{X}_{t_4}^{(u_4)}$ and $\tilde{X}_{t_5}^{(u_5)}$ are made using formula (4.59). For example, the measure $\tilde{X}_{t_5}^{(u_5)}$ projected backward by a timestep h gives the distribution $p\left(X_{t_5-h}^{(u_5)} = . \,|X_{t_5}^{(u_5)} = \tilde{X}_{t_5}^{(u_5)}\right)$, whose variance is again increasing with h. We wish to emphasize two points. First, each of the random processes $X_t^{(u_i)}$ for $i = \{1, 2, 3, 4, 5\}$ is the RCS observed from a different position. If they are normalized by their mean value, such as in Field's model, equations (4.35) and (4.59) are applicable to any of them. Second, if the projection of a measure is not too far forward or backward, the result of the projection is a distribution different from the asymptotic (stationary) distribution, which is that of the unconditioned random variable $X_t^{(u_i)}$ for any i and t. The time series of deterministic measures $\{\tilde{X}_{t_1}^{(u_1)}, \tilde{X}_{t_2}^{(u_2)}, \tilde{X}_{t_3}^{(u_3)}, \tilde{X}_{t_4}^{(u_4)}, \tilde{X}_{t_5}^{(u_5)}\}$ transforms by projection to the reference time t_3 into a series of random variables $\{\hat{X}_{t_3}^{(u_1)}, \hat{X}_{t_3}^{(u_2)}, \hat{X}_{t_3}^{(u_3)}, \hat{X}_{t_3}^{(u_4)}, \hat{X}_{t_3}^{(u_5)}\}$ where $\hat{X}_{t_3}^{(u_i)}$ is the projection of $\tilde{X}_{t_1}^{(u_1)}$ at time t_3. Of course here, $\hat{X}_{t_3}^{(u_3)} = \tilde{X}_{t_3}^{(u_3)}$. The series of deterministic measures from different positions and times transformed into a series of probabilistic measures (random variables) from different positions at the same time. It can be noted that the choice of synchronization time is arbitrary.
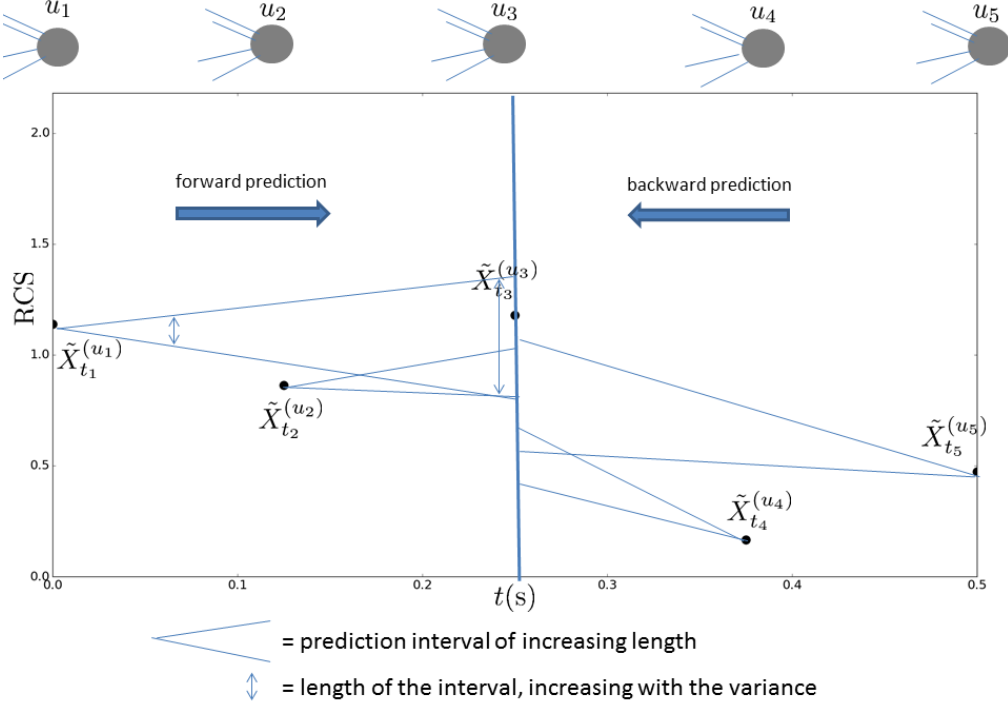
Figure 4.9: Forward and backward inferences of x_t to the common time $t = 0.25\ s$.

4.4.3 Remarks on the Kolmogorov-Smirnov tests

In sections 4.2.1 and 4.2.2, we performed a Kolmogorov-Smirnov test (KS-test) to check the adequation between the analytical and numerical distributions for the transition probabilities of $\gamma_t^{(R)}$ (or $\gamma_t^{(I)}$) and x_t. The test was positive and we concluded that the numerical schemes that we use are precise enough. We did not make the test for the other processes like R_t (or I_t) and z_t for numerical reasons. The test was already computationally demanding for x_t due to the expression of its transition probabilities. The KS-test requires to integrate (numerically) the transition probabilities since it compares cumulative distributions. For R_t for example, it would require an additional integration as evidenced by equation (4.42), which is not realistic. We think that it is sufficient to notice that the KS-test 'validated' the transition probabilities of x_t, $\gamma_t^{(R)}$, $\gamma_t^{(I)}$, and that all the other processes can be derived from them.

4.5 Conclusions

This chapter gives mathematical expressions for the forward and backward transition probabilities of the sea surface speckle, the RCS (texture), the real and imaginary parts of the reflectivity, the intensity, and the complex reflectivity represented in the $(x_t, \gamma_t^{(R)}, \gamma_t^{(I)})$ coordinate system. We solved the Fokker-Planck equations of the speckle and RCS to obtain their transition probabilities: equation (4.12) for the speckle and equations (4.35) and (4.36) for the RCS. From these, we computed the transition probabilities of the real and imaginary parts of the reflectivity and the intensity (formula (4.42) and (4.49)). We also obtained equation (4.53) for the transition probabilities of the complex reflectivity. Numerical simulations systematically reveal an accurate fit between the analytical and numerical distributions,

which were corroborated by Kolmogorov-Smirnov tests. They also illustrate how the initial deterministic measure progressively transforms into an asymptotically distributed random variable with increasing time (see section 4.2). Using the rules of calculus of conditioned probabilities, we reversed the conditioning to obtain backward transition probabilities: formula (4.57), (4.59), (4.63), (4.73) and (4.74).

A series of deterministic measures of the complex reflectivity from different positions and times can then be processed to get a series of probabilistic measures of the speckle, RCS, real (and imaginary) reflectivity, intensity and complex reflectivity from different positions at the same time (synchronization).

All of the formula obtained here depend on three parameters which control the SDE of the RCS and the speckle and have not been estimated yet: $\alpha, \mathcal{A}$ and $\mathcal{B}$. The next chapter is dedicated to the estimation of these parameters by maximum likelihood.

Chapter 5

Estimation of the parameters of Field's model for the sea clutter

We remind that the non-normalized complex reflectivity of the sea clutter is (section 3.3):

$$\Psi_t^{(C)} = C x_t^{1/2} \gamma_t. \tag{5.1}$$

The normalized complex reflectivity ($C = 1$) is (Field's model):

$$\Psi_t = x_t^{1/2} \gamma_t. \tag{5.2}$$

The processes x_t (RCS) and $\gamma_t = \gamma_t^{(R)} + i\gamma_t^{(I)}$ (speckle) are solutions to the stochastic differential equations (SDE):

$$\begin{cases} \mathrm{d}x_t = \mathcal{A}(1 - x_t)\mathrm{d}t + \left(2\frac{\mathcal{A}}{\alpha}x_t\right)^{\frac{1}{2}} \mathrm{d}W_t^{(x)} \\ \mathrm{d}\gamma_t^{(R)} = -\frac{1}{2}\mathcal{B}\gamma_t^{(R)}\mathrm{d}t + \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\mathrm{d}W_t^{(R)} \\ \mathrm{d}\gamma_t^{(I)} = -\frac{1}{2}\mathcal{B}\gamma_t^{(I)}\mathrm{d}t + \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\mathrm{d}W_t^{(I)}. \end{cases} \tag{5.3}$$

Three constants parameterize the equations in (5.3): α, $\mathcal{A}$ for the RCS, and $\mathcal{B}$ for the speckle. We explained in section 4.4 that $\mathcal{A}$ and $\mathcal{B}$ are inverse of the autocorrelation times of x_t and γ_t respectively. We will explain the parameter α in section 5.1. The goal of this chapter is to estimate these parameters.

Many different methods for estimating the parameters of SDE exist in the literature (maximum likelihood, method of moments, etc) [103]. For $\mathcal{A}$ and $\mathcal{B}$, we choose maximum likelihood estimators, and for α, we choose an ergodicity-based estimator (moment). We think that these choices are not completely arbitrary but are adequate given the meaning of the parameters. α is the inverse of the stationary variance of x_t, so using ergodicity to estimate it is obvious and simple. $\mathcal{A}$ and $\mathcal{B}$ parameterize dynamics, so they appear in transition probabilities and therefore in likelihood functions (equation (2.68)). Since we have already computed transition probabilities in the previous chapter for a different purpose, tackling parameter estimation and using maximum likelihood is very natural.

Even if we limit ourselves to maximum likelihood for $\mathcal{A}$ and $\mathcal{B}$, there is some freedom for computing the transition probabilities implied in the likelihood (equation (2.68)). The best option seems to be analytical expressions if they are available, but it might require numerical minimization. A more simplistic approach is to use the Euler approximation, in which case the transition probability is approximated by Gaussian densities from which

it can be very straightforward to maximize the likelihood with respect to the parameter. Between the analytical formula and the Euler approximation, several degrees of complexity exist to approximate the transition probabilities (see [110]).

Is the most sophisticated approach always the best choice? Given an estimator which theoretically performs well, how does this estimator perform if constraints from real world applications are taken into account (such as duration and timesteps of input time series, realistic values of the true parameter etc)?

In this chapter, we also address these two questions. We derive estimators for the parameters $\mathcal{A}, \mathcal{B}, \alpha$, and confront their ability to estimate correctly in relation with the specificities of our application (radar, sea surface). We assess them using numerical simulations. The parameter C is briefly treated for the sake of completeness, though it does not belong to Field's model. For $\mathcal{A}$ and $\mathcal{B}$, different level of approximations are tested for their transition probabilities.

In appendix A.1, we derive volatility-based estimators for $\mathcal{A}$ and $\mathcal{B}$, which show similar performances to the maximum likelihood estimators under the chosen timestep and duration. We have explained in section 2.5 that volatility-based estimation works for high sampling frequency and that it is theoretically not as good as maximum likelihood. The high sampling frequency hypothesis is met in this chapter and in appendix A.1, but it will weaken in chapter 7 and the volatility-based estimators will therefore become inoperative. However, we really recommend to read the appendix after this chapter, in order to contemplate the simplicity with which we can derive volatility-based estimators.

In section 5.1, we estimate the parameters C and α by ergodicity. More estimators from the literature are given for α, since besides being a parameter of Field's model, it is also a parameter of the already known K distribution. Numerical simulations are made to assess the estimators for α. In section 5.2, we estimate $\mathcal{A}$ and $\mathcal{B}$ by ML. We propose and compare three expressions for the transition probabilities: Euler approximation, Nowman's approximation, and the exact closed-form expressions for the transition probabilities. Numerical simulations are also carried to assess the performance of the ML estimators. A joint estimation of $\mathcal{A}$ and α is proposed in section 5.3. Section 5.4 is an extensive discussion whose purpose it to compare the estimators. We compare trajectories of the RCS generated with true parameters and estimated parameters for both Euler's approximation and the exact transition probabilities. We compute root mean square errors after debiasing and show that the estimators have equivalent performances. In section 5.5, we study a possibly new approach for estimating $\mathcal{B}$. Finally, section 5.6 concludes.

5.1   Estimation of C and α

In this section, we assume that the sea clutter is ergodic and use this property to estimate the parameters C and α. This assumption yields simple estimators, but which have the disadvantage of requiring long trajectories compared to what will be necessary for $\mathcal{A}$ and $\mathcal{B}$. It arises as a necessity to have sufficiently many decorrelated samples in the time series, although strictly speaking it is never the case that two samples are decorrelated. Indeed, the autocorrelation of x_t is $corr(x_t, x_{t+\Delta t}) = e^{-\mathcal{A}\Delta t}$. It decays to zero as $\Delta t \to +\infty$, but never reaches it. In practice, we can consider that x_t and $x_{t+\Delta t}$ are decorrelated for large Δt.

5.1.1 Estimation of C

We know from section 3.3 that C^2 is the average power of the reflectivity, i.e.:

$$\mathbb{E}[|\Psi_t^{(C)}|^2] = C^2. \tag{5.4}$$

By ergodicity we have:

$$\lim_{T \to +\infty} \frac{1}{T} \int_0^T |\Psi_t^{(C)}|^2 \mathrm{d}t = \mathbb{E}[|\Psi_0^{(C)}|^2] = C^2. \tag{5.5}$$

We want to estimate C^2 from a discrete time series $\Psi_i^{(C)}$ with i ranging from 0 to n, i.e. $\{\Psi_i^{(C)}, \ i = 0, 1, \ldots, n\}$. From now and until the end of this chapter, we assume a constant timestep Δt. The measurements $\{\Psi_i^{(C)}, \ i = 0, 1, \ldots, n\}$ are therefore made at times $\{t_i, \ i = 0, 1, \ldots, n\}$ with $t_i - t_{i-1} = \Delta t$ for all i. We make the approximation:

$$C^2 = \lim_{T \to +\infty} \frac{1}{T} \int_0^T |\Psi_t^{(C)}|^2 \mathrm{d}t \approx \frac{1}{t_n} \int_0^{t_n} |\Psi_t^{(C)}|^2 \mathrm{d}t \approx \frac{1}{n\Delta t} \sum_{i=0}^{n-1} |\Psi_i^{(C)}|^2 \Delta t. \tag{5.6}$$

Therefore, the ergodicity-based estimator for C (assumed positive) is:

$$\tilde{C} = \left[\frac{1}{n\Delta t} \sum_{i=0}^{n-1} |\Psi_i^{(C)}|^2 \Delta t \right]^{1/2}. \tag{5.7}$$

In practice, given the time series $\{\Psi_i^{(C)}, \ i = 0, 1, \ldots, n\}$, we would compute $\tilde{C}$ and work on the normalized time series $\{\Psi_i^{(C)}/\tilde{C}, \ i = 0, 1, \ldots, n\}$. From now on, we make the approximation that this is the true trajectory, i.e.:

$$\Psi_i^{(C)}/\tilde{C} \approx \Psi_i^{(C)}/C. \tag{5.8}$$

Field's model does indeed assume that $C = 1$. Our purpose in this section was just to show that it is possible in principle to estimate C, and we now go back to Field's model.

5.1.2 Estimation of α

Estimation of α from x_t

According to section 3.3, the stationary distribution of x_t is:

$$p(x_t = x) = \frac{\alpha^\alpha x^{\alpha-1} e^{-\alpha x}}{\Gamma(\alpha)}. \tag{5.9}$$

From the properties of the gamma distribution, it holds that $\forall t$:

$$\begin{cases} \mathbb{E}[x_t] = 1 \\ \mathrm{var}(x_t) = \frac{1}{\alpha}. \end{cases} \tag{5.10}$$

Therefore, α is the inverse of the stationary variance of x_t. We assume that the RCS is ergodic, or more precisely:

$$\lim_{T \to +\infty} \frac{1}{T} \int_0^T (x_t - 1)^2 \mathrm{d}t = \mathbb{E}[(x_0 - 1)^2] = \frac{1}{\alpha}, \tag{5.11}$$

To compute α from a discrete time series x_i with i ranging from 0 to n and constant timestep Δt, we make the approximation:

$$\frac{1}{\alpha} = \lim_{T \to +\infty} \frac{1}{T} \int_0^T (x_t - 1)^2 \mathrm{d}t \approx \frac{1}{t_n} \int_0^{t_n} (x_t - 1)^2 \mathrm{d}t \approx \frac{1}{n\Delta t} \sum_{i=0}^{n-1} (x_i - 1)^2 \Delta t. \qquad (5.12)$$

Therefore, the ergodicity-based estimator for α (from x_t) is:

$$\tilde{\alpha}_x = \left[\frac{1}{n\Delta t} \sum_{i=0}^{n-1} (x_i - 1)^2 \Delta t \right]^{-1}. \qquad (5.13)$$

Estimation of α from z_t

We have seen in section 3.3 that in Field's model, for all t the intensity $z_t = |\Psi_t|^2$ is K-distributed. More precisely,

$$p(z_t = x) = \frac{2b^{(\nu+1)/2} x^{(\nu-1)/2}}{\Gamma(\nu)} K_{\nu-1}(2\sqrt{bx}), \qquad (5.14)$$

with $b = \nu = \alpha$. It is known that for the K distribution, the n-th moment is given by (see [141] p 110):

$$\frac{\langle z_t^n \rangle}{\langle z_t \rangle^n} = n! \frac{\Gamma(n+\nu)}{\Gamma(\nu)\nu^n}. \qquad (5.15)$$

where $\langle . \rangle$ is another notation for $\mathbb{E}[.]$. For $n = 2$, we have:

$$\langle z_t^2 \rangle = \langle z_t \rangle^2 2 \frac{\Gamma(2+\nu)}{\Gamma(\nu)\nu^2} = \langle z_t \rangle^2 \frac{\nu(\nu+1)}{\nu^2} = 2\langle z_t \rangle^2 \left(1 + \frac{1}{\nu} \right). \qquad (5.16)$$

Therefore,

$$\mathrm{var}(z_t) = \langle z_t^2 \rangle - \langle z_t \rangle^2 = \langle z_t \rangle^2 \left(1 + \frac{2}{\nu} \right). \qquad (5.17)$$

If we isolate $\nu = \alpha$, we get:

$$\alpha = \frac{2\langle z_t \rangle^2}{\mathrm{var}(z_t) - \langle z_t \rangle^2}. \qquad (5.18)$$

By the hypothesis of ergodicity, we assume that:

$$\langle z_t \rangle = \lim_{T \to +\infty} \frac{1}{T} \int_0^T z_t \mathrm{d}t \approx \frac{1}{n\Delta t} \sum_{i=0}^{n-1} z_i \Delta t \qquad (5.19)$$

$$\mathrm{var}(z_t) = \lim_{T \to +\infty} \frac{1}{T} \int_0^T (z_t - \langle z_t \rangle)^2 \mathrm{d}t \approx \frac{1}{n\Delta t} \sum_{i=0}^{n-1} (z_i - 1)^2 \Delta t. \qquad (5.20)$$

The estimator for α from z_t is:

$$\tilde{\alpha}_z = \frac{2 \left(\sum_{i=0}^{n-1} z_i \Delta t \right)^2}{\sum_{i=0}^{n-1} (z_i - 1)^2 \Delta t - \left(\sum_{i=0}^{n-1} z_i \Delta t \right)^2}. \qquad (5.21)$$

Over the estimator given by equation (5.13), $\tilde{\alpha}_z$ has the advantage that it does not require x_t. In practice, x_t is not directly observed but must be estimated, unlike z_t which is actually

observed. For the problem of the observability of x_t, see chapter 7. The moment-based estimator $\tilde{\alpha}_z$ has been reported in [118] section 5.6, as well as in [141] chapter 9 equation (9.46). In [141], it also accounts for noise. We cite two more estimators presented in [118]. First, it is possible to define a maximum-likelihood estimator for α. If we observe $z = \{z_0, z_1, \ldots, z_n\}$ at times $t_0 < t_1 < \cdots < t_n$, and if we assume that the z_{t_i} are independent for different i, then the likelihood of z as a function of α is:

$$\mathcal{L}(z, \alpha) = \prod_{i=0}^{n} \frac{2\alpha^{(\alpha+1)/2} z_i^{(\alpha-1)/2}}{\Gamma(\alpha)} K_{\alpha-1}(2\sqrt{\alpha z_i}). \tag{5.22}$$

By numerical maximization of $\mathcal{L}(z, \alpha)$ with respect to α, we obtain the ML estimator of α, denoted $\tilde{\alpha}_{ML}$.

Another estimator is given further in section 5.7 of [118]: the normalized logarithm estimator. We denote it $\tilde{\alpha}_{log}$. It is explained in [118] that it satisfies the implicit equation:

$$\ln\langle z \rangle - \langle \ln z \rangle = \ln \tilde{\alpha}_{log} - \Psi^{(0)}(\tilde{\alpha}_{log}) + \ln L - \Psi^{(0)}(L). \tag{5.23}$$

$\Psi^{(0)}$ is the digamma function, and L is a parameter of a more general (than ours) expression of the K distribution proposed in [118]. By correspondence, we have $L = 1$. Equation (5.23) must be solved numerically to get $\tilde{\alpha}_{log}$.

Numerical experiments

To assess the estimators of α ($\tilde{\alpha}_x$, $\tilde{\alpha}_z$, $\tilde{\alpha}_{ML}$ and $\tilde{\alpha}_{log}$), we do numerical simulations. α is dimensionless and we know that its value typically ranges from 0.1 to $+\infty$. This range can be found in [141] p 110-111. Indeed, α being the shape parameter of the K distribution, it has already been estimated from real data. We choose to explore the range $\alpha \in [0.1, 10]$. The performance of $\tilde{\alpha}_x$ (based on x_t) are much better than those of $\tilde{\alpha}_z$, $\tilde{\alpha}_{ML}$ and $\tilde{\alpha}_{log}$ (based on z_t). For that reason, we present the performances of $\tilde{\alpha}_x$ first and separately.

For each value of α in $\{0.1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, we simulate $N = 1000$ trajectories of x_t. To do so, we solve numerically the SDE in equation (5.3) using Milstein's scheme for x_t. See definition 2.38 for Milstein's scheme, and [62] for an introduction to numerical simulation of SDE. The more simple Euler-Maruyama scheme is not used for x_t because it raises numerical issues. Indeed, it could generate negative values, which is absurd since the RCS is always positive. That Milstein's scheme is more performant stems from the fact that is has a strong order of convergence of 1, while Euler-Maruyama's scheme has a strong order of convergence of 1/2 [62].

In our experiments, the simulation timestep, $\hat{\Delta}t$, should ideally be smaller than the measurement timestep Δt quoted above. Consequently, the generated trajectories are evaluated at times $\hat{t}_k$ for $k = 0...mn$, where m is the decimation ratio. Then, for the estimation stage, they are downsampled (i.e decimated) to the times t_i for $i = 0...n$, with $t_i = \hat{t}_{mi}$. A realistic value for Δt is $0.001\,\mathrm{s}$ since the Pulse Repetition Frequency is about $1\,\mathrm{kHz}$ for satellite and airborne applications [35]. Consequently, we choose $\hat{\Delta}t = 10^{-5}\,\mathrm{s}$ leading to a decimation ratio $m = 100$. Note that we have verified that $\hat{\Delta}t = 10^{-5}\,\mathrm{s}$ is small enough: changing it to $\hat{\Delta}t = 10^{-7}\,\mathrm{s}$ does not alter the results presented here but it requires a much longer computing time.

For every of the N simulated trajectories at fixed α, we obtain N estimates $\tilde{\alpha}_1, \tilde{\alpha}_2, ..., \tilde{\alpha}_N$ from the downsampled trajectories (i.e. trajectories with timestep equal to $0.001\,\mathrm{s}$), where

$\tilde{\alpha}$ is either of the estimators $\tilde{\alpha}_x$, $\tilde{\alpha}_z$, $\tilde{\alpha}_{ML}$ and $\tilde{\alpha}_{log}$. Then, the estimation bias $b(\tilde{\alpha})$ and (unbiased) standard deviation $\sigma(\tilde{\alpha})$ are computed:

$$\tilde{b}(\alpha) = \frac{1}{N} \sum_{i=1}^{N} (\tilde{\alpha}_i - \bar{\alpha}), \tag{5.24}$$

$$\tilde{\sigma}^2(\alpha) = \frac{1}{N-1} \sum_{i=1}^{N} (\tilde{\alpha}_i - \bar{\alpha})^2. \tag{5.25}$$

Figure 5.1 represents the estimation bias and standard deviation of $\tilde{\alpha}_x$ for 1000 trajectories for different durations of the trajectories and values of $\mathcal{A}$. We see that even with 300 s, the standard deviation is substantial compared to the true value of α. The estimation of α requires a much longer observation of the sea surface than the estimation of $\mathcal{A}$ (see section 5.2), due to its different physical meaning and mathematical expression. $\mathcal{A}$ is the inverse of a decorrelation time and therefore a time series $1/\mathcal{A}$ s long (order of magnitude) is sufficient to have a satisfying estimation of $\mathcal{A}$. α is the variance of x_t for any t. To estimate it satisfactorily, one ought to average enough independent realizations of x_t, which is approximated using the ergodic hypothesis, *i.e.* by averaging over one trajectory. However, two values $\tilde{x}_{t_1}$ and $\tilde{x}_{t_2}$ of one trajectory constitute roughly two independent realizations of the same random variable only if $t_2 - t_1$ is large enough. The correlation between x_{t_1} and x_{t_2} can be shown to be $corr(x_{t_1}, x_{t_2}) = e^{-\mathcal{A}(t_2-t_1)}$, which gives a correlation of 0.05 if $\mathcal{A}(t_2-t_1) = 3$, in which case we can assume that x_{t_1} and x_{t_2} are independent. Figure 5.1 illustrates the fact that the standard deviation and bias of the estimators look alike for $\mathcal{A} = 1\,\text{Hz}$, $T = 300\,\text{s}$, and $\mathcal{A} = 10\,\text{Hz}$, $T = 30\,\text{s}$. This makes sense since both these configurations give approximately 100 independent realizations of the same random variable. The resemblance is even more striking for $\mathcal{A} = 1\,\text{Hz}$, $T = 30\,\text{s}$, and $\mathcal{A} = 10\,\text{Hz}$, $T = 3\,\text{s}$.

Numerical simulations showed us that $\tilde{\alpha}_z$, $\tilde{\alpha}_{ML}$ and $\tilde{\alpha}_{log}$ are not performant at all for durations as short as 300 s (and $\mathcal{A} = 1$ Hz). In particular, they generate a lot of extreme values, or even negative ones. We choose a duration of 1200 s to assess them. Due to computer limitations, we set $\hat{\Delta}t = \Delta t = 10^{-3}$ s. For each value of α in $\{0.1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, we simulate 1000 trajectories of x_t and γ_t (using Milstein's scheme for x_t and Euler-Maruyama's scheme for γ_t) from which we compute $z_t = x_t |\gamma_t|^2$, from which $\tilde{\alpha}_z$, $\tilde{\alpha}_{ML}$ and $\tilde{\alpha}_{log}$ are computed. For comparison, $\tilde{\alpha}_x$ is also computed. $\tilde{\alpha}_{ML}$ and $\tilde{\alpha}_{log}$ are found by numerical search (see section 5.1.2) over the space $\{0.01, 0.02, 0.03, \ldots, 14.99, 15\}$. For all trajectories, we set $\mathcal{A} = 1$ Hz and $\mathcal{B} = 100$ Hz.

The estimation bias and standard deviation as a function of true α are represented in figure 5.2. There is a very clear hierarchy, especially for $\alpha > 5$. From the best to the worst estimator for α, we have: $\tilde{\alpha}_x$, $\tilde{\alpha}_z$, $\tilde{\alpha}_{ML}$ and then $\tilde{\alpha}_{log}$. For low α, the performance are similar, though $\tilde{\alpha}_{ML}$ is very biased for $\alpha = 0.1$. The bias of $\tilde{\alpha}_x$ and $\tilde{\alpha}_z$ are low, and grow linearly as α increases, suggesting very little statistical effect. It reaches 0.08 at $\alpha = 10$ for $\tilde{\alpha}_z$, which may be negligible compared to the standard deviation at that same value. For $\tilde{\alpha}_{ML}$ and $\tilde{\alpha}_{log}$, the bias is much larger and some large value are reached occasionally, suggesting a non-negligible statistical effect. Figure 5.3 represents the first 50 estimations of α with the four estimators. Though some correlation between the estimators seems to exist, it is not absolutely clear.

We will use again the estimation of α in chapter 7. Only $\tilde{\alpha}_x$ and $\tilde{\alpha}_z$ shall be kept from now, in particular because $\tilde{\alpha}_z$ is the best of the z_t-based estimators of α, while $\tilde{\alpha}_x$ is the only x_t-based estimator of α, and the best overall.
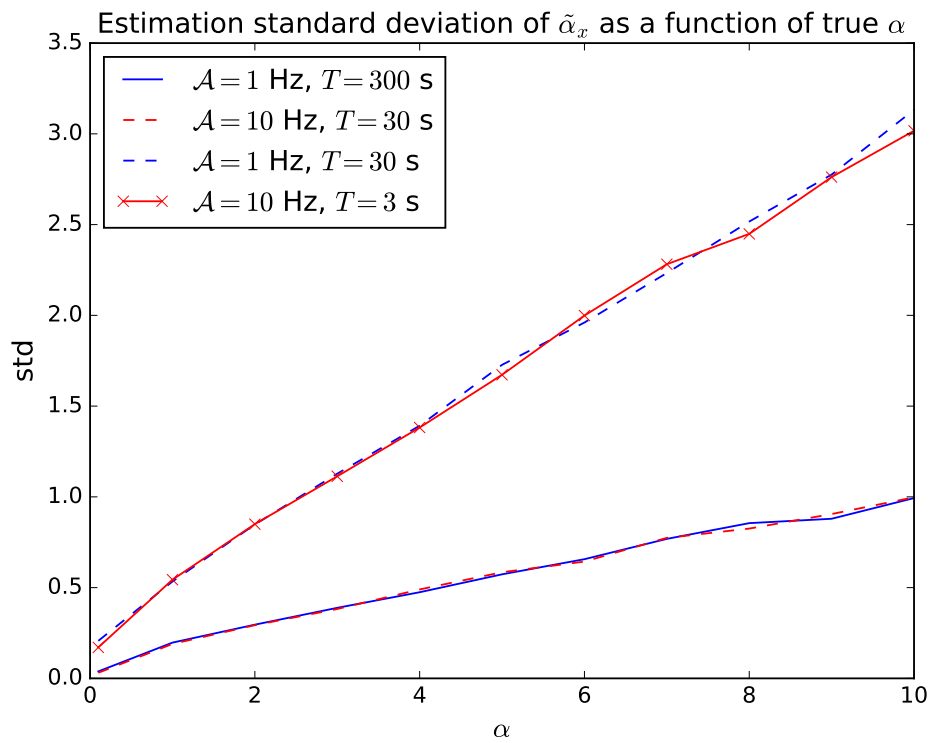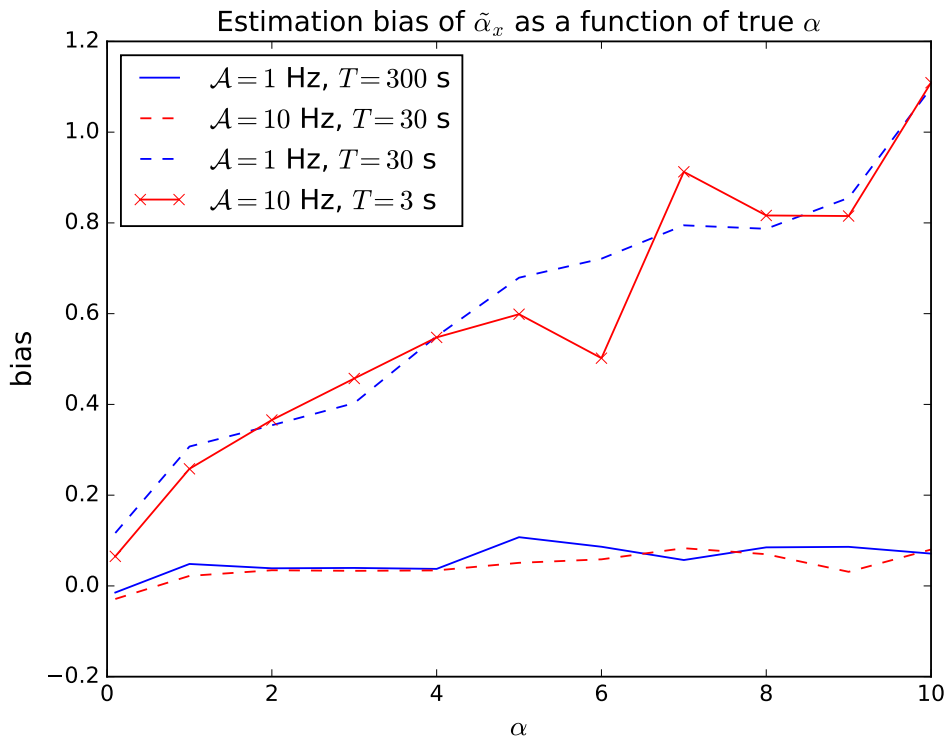
Figure 5.1: Estimation bias (up) and standard deviation (down) of $\tilde{\alpha}_x$ as a function of true α based on 1000 trajectories for $T = 300$ s.
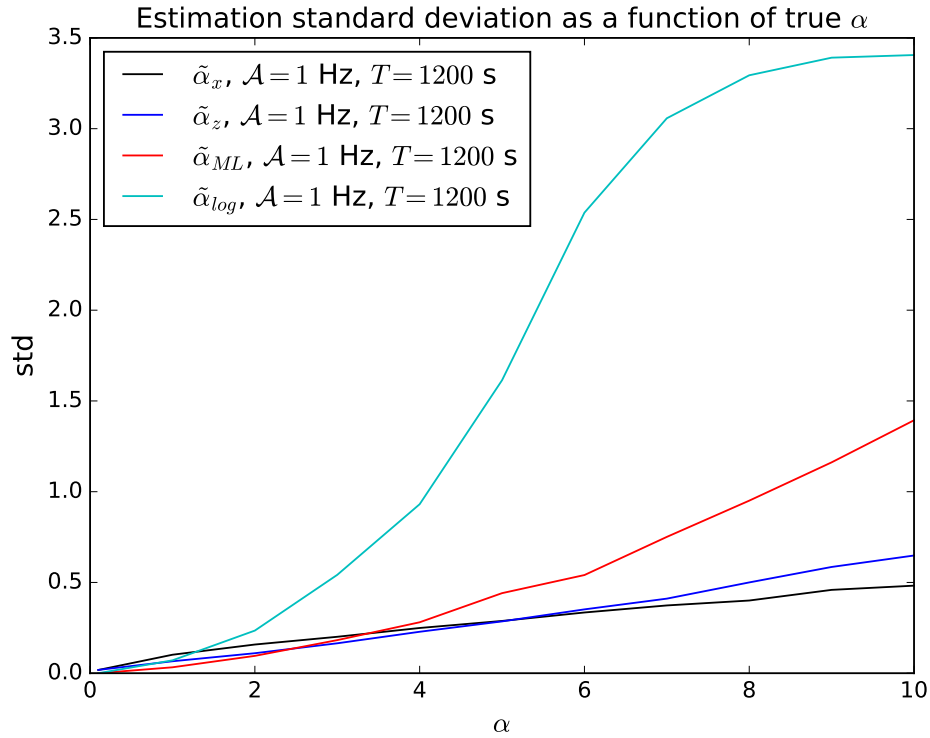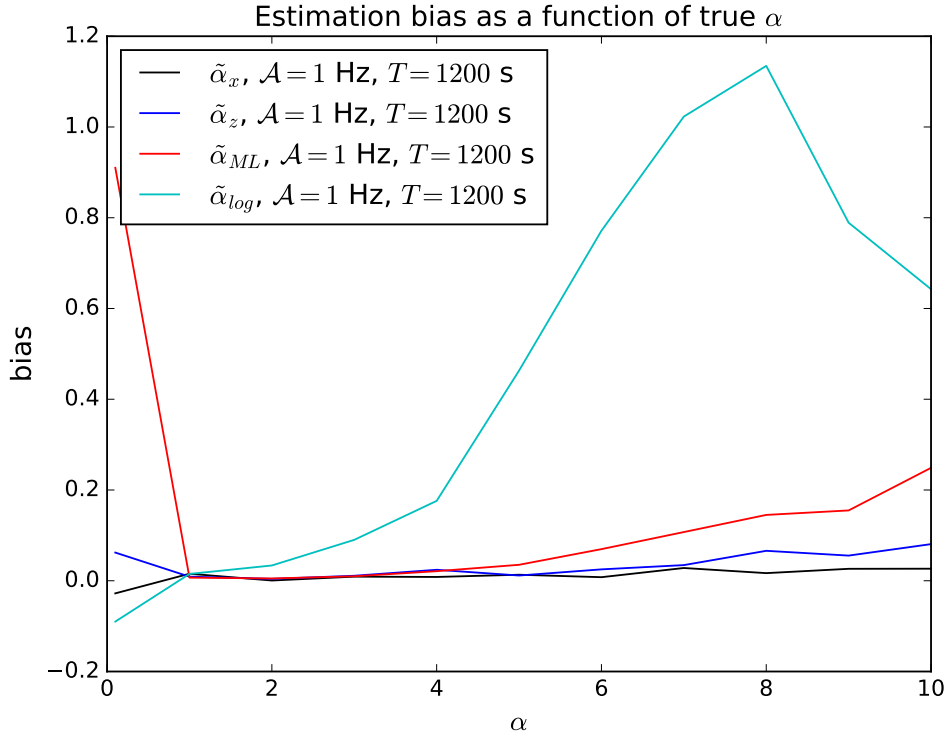
Figure 5.2: Estimation bias (up) and standard deviation (down) of $\tilde{\alpha}_x$, $\tilde{\alpha}_z$, $\tilde{\alpha}_{ML}$ and $\tilde{\alpha}_{log}$ as a function of true α based on 1000 trajectories of duration 1200 s each, with $\mathcal{A} = 1$ Hz and $\mathcal{B} = 100$ Hz.
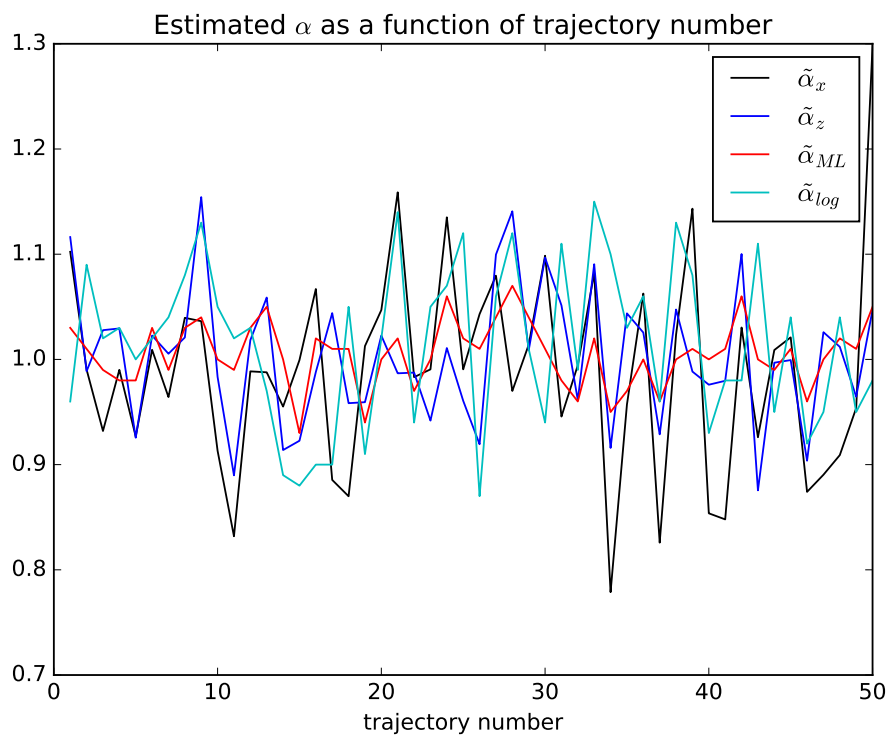
Figure 5.3: Comparison between $\tilde{\alpha}_x$, $\tilde{\alpha}_z$, $\tilde{\alpha}_{ML}$ and $\tilde{\alpha}_{log}$ as a function of trajectory number, for the first 50 trajectories.

5.2 Maximum Likelihood estimation of $\mathcal{A}$ and $\mathcal{B}$

In this section, we propose to estimate $\mathcal{A}$ and $\mathcal{B}$ by maximum likelihood. From radar data, one can already note that the speckle variation timescale is $10\,\text{ms}$ while the RCS variation timescale is about $1\,\text{s}$ [140], [43]. Therefore, the corresponding orders of magnitude of $\mathcal{A}$ and $\mathcal{B}$ are $\mathcal{A} = 1\,\text{Hz}$ and $\mathcal{B} = 100\,\text{Hz}$. In a real data situation, we would have access to the complex-valued time series $C\tilde{\Psi}_t$. As already stated, by ergodicity estimation of C we consider that we start from the normalized reflectivity $\tilde{\Psi}_t$. Using the difference between the slow dynamics of x_t and fast dynamics of $\gamma_t^{(R)}$, $\gamma_t^{(I)}$, one can retrieve x_t (see [44]) and then $\gamma_t^{(R)}$ and $\gamma_t^{(I)}$. We therefore assume that we observe three discrete time series: $\tilde{x}_i, \tilde{\gamma}_i^{(R)}, \tilde{\gamma}_i^{(I)}$ where the measurements have been made at times t_i with i ranging from 0 to n with constant timestep Δt. We notice that $\gamma_t^{(R)}$ and $\gamma_t^{(I)}$ follow the same SDE. Since they are independent, it is sufficient to use $\tilde{\gamma}_i^{(R)}$ to estimate $\mathcal{B}$. Alternatively, we can average the two estimates arising from these two time series.

We remind that the ML estimation consists in maximizing the likelihood function with respect to the parameters, *i.e.* determining the parameters that makes the observed data the most probable. Because the process x_t and $\gamma_t^{(R)}$ are Markov processes, the likelihood function for the RCS is (see section 2.3.1):

$$\mathcal{L}(\mathcal{A}, \alpha) = p(x_{t_0} = \tilde{x}_0) \prod_{i=1}^{n} p(x_{t_i} = \tilde{x}_i \mid x_{t_{i-1}} = \tilde{x}_{i-1}), \qquad (5.26)$$

and the likelihood function for the speckle is:

$$\mathcal{L}(\mathcal{B}) = p\left(\gamma_{t_0}^{(R)} = \tilde{\gamma}_0^{(R)}\right) \prod_{i=1}^{n} p\left(\gamma_{t_i}^{(R)} = \tilde{\gamma}_i^{(R)} \mid \gamma_{t_{i-1}}^{(R)} = \tilde{\gamma}_{i-1}^{(R)}\right). \qquad (5.27)$$

We will assume that α is known. The optimality conditions (see section 2.3.1 equation (2.70)) read:

$$\frac{\partial \ln \mathcal{L}}{\partial \mathcal{A}}(\tilde{\mathcal{A}}, \alpha) = 0 \qquad (5.28)$$

$$\frac{\partial \ln \mathcal{L}}{\partial \mathcal{B}}(\tilde{\mathcal{B}}) = 0. \qquad (5.29)$$

The only difference between equations (2.70) and (5.28) we take the log before maximizing in (5.28), which is equivalent since the log is strictly increasing. In this section, we compare three different models for the transition probabilities of x_t and $\gamma_t^{(R)}$: the Euler approximation, Nowman's approximation and the exact transition probabilities.

5.2.1 Euler's approximation

Euler's approximation relies on the discretization method of Euler-Maruyama, which enables both to simulate numerical trajectories (*i.e.* solve numerically the SDE) and to have an analytical discrete-time approximation of our continuous-time process. In Euler's approximation, the drift and volatility of the SDE are assumed constant over the interval $[t_{i-1}, t_i]$ (see definition 2.37). Applied to the SDE of x_t, *i.e.* the first equation of (5.3), and assuming a constant timestep Δt, Euler-Maruyama's scheme gives:

$$x_{t_i} \approx x_{t_{i-1}} + \mathcal{A}(1 - x_{t_{i-1}})\Delta t + \left(\frac{2\mathcal{A}}{\alpha} x_{t_{i-1}}\right)^{\frac{1}{2}} \mathcal{N}(0, \Delta t). \qquad (5.30)$$

Applying Euler-Maruyama's scheme to the second equation of (5.3), we get:

$$\gamma_{t_i}^{(R)} \approx \gamma_{t_{i-1}}^{(R)} - \frac{1}{2}\mathcal{B}\gamma_{t_{i-1}}^{(R)}\Delta t + \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\mathcal{N}(0, \Delta t). \tag{5.31}$$

From equations (5.30) and (5.31), we get the following transition probabilities:

$$\begin{cases} p(x_{t_i} = \tilde{x}_i \mid x_{t_{i-1}} = \tilde{x}_{i-1}) = \frac{\sqrt{\alpha}}{\sqrt{4\pi\tilde{x}_{i-1}\mathcal{A}\Delta t}}e^{-\frac{\alpha(\tilde{x}_i - \mathcal{A}\Delta t - (1-\mathcal{A}\Delta t)\tilde{x}_{i-1})^2}{4\mathcal{A}\Delta t\tilde{x}_{i-1}}}, \\ p\left(\gamma_{t_i}^{(R)} = \tilde{\gamma}_i^{(R)} \mid \gamma_{t_{i-1}}^{(R)} = \tilde{\gamma}_{i-1}^{(R)}\right) = \frac{1}{\sqrt{\pi\mathcal{B}\Delta t}}e^{-\frac{\left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)}(1-\mathcal{B}\Delta t/2)\right)^2}{\mathcal{B}\Delta t}}. \end{cases} \tag{5.32}$$

We inject these expressions in the corresponding likelihood functions (5.26) and (5.27). Using (5.26) and assuming that we know α, we express the condition $\frac{\partial \ln \mathcal{L}}{\partial \mathcal{A}}(\tilde{\mathcal{A}}, \alpha) = 0$ to estimate $\mathcal{A}$. We also express the condition $\frac{\partial \ln \mathcal{L}}{\partial \mathcal{B}}(\tilde{\mathcal{B}}) = 0$. $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{B}}$ are the values of $\mathcal{A}$ and $\mathcal{B}$ which maximize the likelihood. After some calculations, it is possible to show that $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{B}}$ are the roots of two second-order polynomials (see appendix C):

$$\begin{cases} -\sum_{i=1}^n \frac{\alpha(\tilde{x}_{i-1}-1)^2}{4\tilde{x}_{i-1}}\mathcal{A}^2 - \frac{n}{2}\mathcal{A} + \sum_{i=1}^n \frac{\alpha(\tilde{x}_i - \tilde{x}_{i-1})^2}{4\tilde{x}_{i-1}\Delta t} = 0 \\ -\sum_{i=1}^n \frac{\tilde{\gamma}_{i-1}^{(R)\,2}\Delta t}{4}\mathcal{B}^2 - \frac{n}{2}\mathcal{B} + \sum_{i=1}^n \frac{\left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)}\right)^2}{\Delta t} = 0. \end{cases} \tag{5.33}$$

For both $\mathcal{A}$ and $\mathcal{B}$, the discriminant of the polynomial is always positive and there is only one positive root (no ambiguity). $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{B}}$ are therefore easily found analytically without numerical minimization of the likelihood function.

5.2.2 Nowman's approximation

Nowman's approach [104, 110] is applicable to SDE with a linear drift, *i.e.* of the form:

$$dX_t = \kappa(\mu - X(t))dt + \sigma(X_t)dW_t, \tag{5.34}$$

where κ and μ are two constants. This is precisely the case for the SDE of x_t and $\gamma_t^{(R)}$, hence our interest in this approach. In the case of a linear drift, the solution to the SDE (5.34) can be expressed as follows:

$$X_{t_i} = X_{t_{i-1}}e^{-\kappa\Delta t} + \mu(1 - e^{-\kappa\Delta t}) + \int_{t_{i-1}}^{t_i} \sigma(X_s)e^{-\kappa(t_i-s)}dW_s. \tag{5.35}$$

Nowman assumes that the volatility is constant over the time interval: $\forall s \in [t_{i-1}, t_i]$ $\sigma(X_s) = \sigma(X_{t_{i-1}})$, in which case:

$$X_{t_i} = X_{t_{i-1}}e^{-\kappa\Delta t} + \mu(1 - e^{-\kappa\Delta t}) + \eta_{t_i}, \tag{5.36}$$

with:

$$\eta_{t_i} = \sigma(X_{t_{i-1}}) \int_{t_{i-1}}^{t_i} e^{-\kappa(t_i-s)}dW_s. \tag{5.37}$$

We can show that $\mathbb{E}(\eta_{t_i}) = 0$ and $\mathbb{E}(\eta_{t_i}^2) = \frac{\sigma(X_{t_{i-1}})^2\left(1-e^{-2\kappa\Delta t}\right)}{2\kappa}$ from which we make the approximation:

$$\eta_{t_i} \sim \mathcal{N}\left(0, \frac{\sigma(X_{t_{i-1}})^2\left(1 - e^{-2\kappa\Delta t}\right)}{2\kappa}\right). \tag{5.38}$$

Combining (5.36) and (5.38), we get:

$$X_{t_i} = X_{t_{i-1}} e^{-\kappa \Delta t} + \mu(1 - e^{-\kappa \Delta t}) + \mathcal{N}\left(0, \frac{\sigma(X_{t_{i-1}})^2 \left(1 - e^{-2\kappa \Delta t}\right)}{2\kappa}\right). \tag{5.39}$$

The transition probabilities are again Gaussian but the drift has been completely resolved. For x_t and $\gamma_t^{(R)}$, we get:

$$\begin{cases} p(x_{t_i} = \tilde{x}_i \mid x_{t_{i-1}} = \tilde{x}_{i-1}) = \dfrac{\sqrt{\alpha}}{\sqrt{2\pi \tilde{x}_{i-1}(1 - e^{-2\mathcal{A}\Delta t})}} e^{-\frac{1}{2} \frac{\alpha\left(\tilde{x}_i - 1 + e^{-\mathcal{A}\Delta t}(1 - \tilde{x}_{i-1})\right)^2}{\tilde{x}_{i-1}(1 - e^{-2\mathcal{A}\Delta t})}} \\[2em] p\left(\gamma_{t_i}^{(R)} = \tilde{\gamma}_i^{(R)} \mid \gamma_{t_{i-1}}^{(R)} = \tilde{\gamma}_{i-1}^{(R)}\right) = \dfrac{1}{\sqrt{\pi(1 - e^{-\mathcal{B}\Delta t})}} e^{-\frac{\left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)} e^{-\mathcal{B}\Delta t/2}\right)^2}{1 - e^{-\mathcal{B}\Delta t}}}. \end{cases} \tag{5.40}$$

Then, we inject these expressions in the corresponding likelihood functions (5.26) and (5.27). The conditions $\frac{\partial \ln \mathcal{L}}{\partial \mathcal{A}}(\tilde{\mathcal{A}}, \alpha) = 0$ and $\frac{\partial \ln \mathcal{L}}{\partial \mathcal{B}}(\tilde{\mathcal{B}}) = 0$ lead to third-order polynomials depending on $X = e^{-\tilde{\mathcal{A}}\Delta t}$ and $Y = e^{-\tilde{\mathcal{B}}\Delta t/2}$:

$$\begin{cases} nX^3 - \sum_{i=1}^n \frac{\alpha(\tilde{x}_i - 1)(\tilde{x}_{i-1} - 1)}{\tilde{x}_{i-1}} X^2 + \left(-n + \sum_{i=1}^n \frac{\alpha(\tilde{x}_i - 1)^2 + \alpha(\tilde{x}_{i-1} - 1)^2}{\tilde{x}_{i-1}}\right) X \\ \qquad\qquad - \sum_{i=1}^n \frac{\alpha(\tilde{x}_i - 1)(\tilde{x}_{i-1} - 1)}{\tilde{x}_{i-1}} = 0 \\[1em] \frac{n}{2}Y^3 - \sum_{i=1}^n \tilde{\gamma}_{i-1}^{(R)} \tilde{\gamma}_i^{(R)} Y^2 + \left(-\frac{n}{2} + \sum_{i=1}^n \left(\tilde{\gamma}_i^{(R)\,2} + \tilde{\gamma}_{i-1}^{(R)\,2}\right)\right) Y - \sum_{i=1}^n \tilde{\gamma}_{i-1}^{(R)} \tilde{\gamma}_i^{(R)} = 0. \end{cases}$$

There is only one real root for each of these two polynomials (the other two being complex conjugates). Thus $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{B}}$ are determined analytically without ambiguity nor the need for numerical minimization. In this regard, Nowman's approximation is very similar to Euler's approximation, but it remains formally better since the drift is exactly solved.

5.2.3 Exact transition probabilities

The transition probabilities of x_t and $\gamma_t^{(R)}$ can be obtained by solving their respective Fokker-Planck equations (see [120]). $\gamma_t^{(R)}$ is an Ornstein-Uhlenbeck process and its Fokker-Planck equation is easily solved by Fourier transform and method of characteristics ([120], chapter 5). In the particular case of $\gamma_t^{(R)}$, it gives (see chapter 4):

$$p\left(\gamma_{t+\Delta t}^{(R)} = x \mid \gamma_t^{(R)} = y\right) = \frac{1}{\sqrt{2\pi v(\Delta t)}} e^{-\frac{1}{2} \frac{(x - m(\Delta t))^2}{v(\Delta t)}}, \tag{5.41}$$

which is Gaussian with expectation

$$m(\Delta t) = y e^{-\mathcal{B}\Delta t/2} \tag{5.42}$$

and variance

$$v(\Delta t) = \frac{1 - e^{-\mathcal{B}\Delta t}}{2}. \tag{5.43}$$

Replacing x, y, t, $t + \Delta t$ by $\tilde{\gamma}_i^{(R)}$, $\tilde{\gamma}_{i-1}^{(R)}$, t_{i-1}, t_i respectively, we obtain the same expression as the second equation of (5.40). This is not surprising since Nowman's method resolves the drift and then assumes locally a constant volatility. From (5.3), we see that the constant volatility assumption is actually always true for $\gamma_t^{(R)}$.

x_t is a Cox-Ingersoll-Ross process. It has been shown that its transition probabilities can be written analytically ([46] and [48] p 63):

$$p(x_{t+\Delta t} = x | x_t = y) = c e^{-cx-c\delta y} \left(\frac{x}{y\delta} \right)^{\frac{\alpha-1}{2}} I_{\alpha-1}(2c\sqrt{xy\delta}), \qquad (5.44)$$

with $c = \frac{\alpha}{1-e^{-\mathcal{A}\Delta t}}$ and $\delta = e^{-\mathcal{A}\Delta t}$ and where $I_{\alpha-1}$ is the modified Bessel function of the first kind of order $\alpha - 1$.

Maximization of the likelihood function is achieved numerically by gradient descent applied to $-\ln(\mathcal{L}(\mathcal{A}, \alpha))$, α being fixed.

5.2.4 Numerical experiments for $\mathcal{A}$ and $\mathcal{B}$

To assess the ability of a method to estimate the parameters, we simulate many trajectories of x_t and $\gamma_t^{(R)}$ with known parameters and then try to retrieve them. To do so, we solve numerically the SDE in equation (5.3) using Euler-Maruyama's scheme for $\gamma_t^{(R)}$, $\gamma_t^{(I)}$ and Milstein's scheme for x_t (see section 2.2.3). The remarks made for α about the simulation timestep $\hat{\Delta} t$ and 'observed' timestep Δt still hold: the trajectories are simulated with $\hat{\Delta} t = 10^{-5}$ s and sampled at $\Delta t = 10^{-3}$ s before estimation. To estimate $\mathcal{A}$ and $\mathcal{B}$ (with order of magnitude around $1\,\text{Hz}$ and $100\,\text{Hz}$ respectively), a duration of $1\,\text{s}$ is sufficient leading to $t_n = t_{1000} = 1\,\text{s}$. The estimations of $\mathcal{A}$ and $\mathcal{B}$ are completely independent since $\mathcal{A}$ requires only the trajectory of the RCS and $\mathcal{B}$ requires only the trajectory of the speckle. However, the estimation of $\mathcal{A}$ does require α, and for simplicity we set its value to 1. We remind that the estimation of α was in section 5.1.2, and that we observed that long trajectories were necessary to estimate it correctly.

Simulations for $\mathcal{A}$

For the numerical simulations, the true values of $\mathcal{A}$ range in the interval $[0.1, 10]\,\text{Hz}$. For each value of $\mathcal{A}$ in $\{0.1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, we generate $N = 1000$ trajectories of the RCS (with $\alpha = 1$) of duration $1\,\text{s}$. We obtain N estimates $\tilde{\mathcal{A}}_1, \tilde{\mathcal{A}}_2, \ldots, \tilde{\mathcal{A}}_N$ from which the estimation bias $\tilde{b}(\mathcal{A})$ and (unbiased) standard deviation $\tilde{\sigma}(\mathcal{A})$ are computed.

Figure 5.4 represents the estimation bias and standard deviation in this configuration, as a function of the true value of $\mathcal{A}$ used to generate the trajectories. The Euler and the Nowman approximations have very similar results, with no improvement when using Nowman's approximation. On the contrary, a slight increase in bias is observed compared to Euler's approximation. For both of them, there is an exponential increase in the bias and standard deviation, which persist if we compute the relative bias and standard deviation (*i.e.* $\tilde{b}(\mathcal{A})/\mathcal{A}$ and $\tilde{\sigma}(\mathcal{A})/\mathcal{A}$). However, using the exact transition probabilities completely annihilates the bias. Even though the standard deviation still increases exponentially, this is only due to the log scale. The relative variance is constant.

Figure 5.4: Estimation bias (up) and standard deviation (down) of $\mathcal{A}$ as a function of true $\mathcal{A}$ based on 1000 trajectories of duration 1 s each. 3 methods are compared: the Euler approximation, the Nowman approximation, and the exact transition probability.

Simulations for $\mathcal{B}$

For the numerical simulations, the true values of $\mathcal{B}$ range in the interval $[10, 1000]$ Hz. For each $\mathcal{B}$ in $\{10, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000\}$ Hz, we generate $N = 1000$ trajectories of $\gamma_t^{(R)}$ of duration 1 s. We obtain N estimates $\tilde{\mathcal{B}}_1, \tilde{\mathcal{B}}_2, \ldots, \tilde{\mathcal{B}}_N$ from which we compute the estimation bias $\tilde{b}(\mathcal{B})$ and standard deviation $\tilde{\sigma}(\mathcal{B})$.

Figure 5.5 represents the estimation bias and standard deviations in this configuration, as a function of the true value of $\mathcal{B}$ used to generate the trajectories. The Euler approximation results in an exponential increase of the bias and standard deviation with increasing $\mathcal{B}$, which holds for the relative bias and standard deviation. However, using the Nowman approximation, which is the exact transition probabilities for this case, completely annihilates the bias. Even though the standard deviation still increases exponentially, this is mostly due to the log scale. The relative variance increases only slightly with increasing $\mathcal{B}$.

5.2.5   Notational remarks

In section 5.1, for the estimators of C and α, we made use of a *time series*, for example $\{x_i, i = 1, 2, \ldots, n\}$ (RCS). $\tilde{\alpha}_x$ was then expressed as a function of it. Since the time series is random, $\tilde{\alpha}_x$ is also random: it is an *estimator*. However, in section 5.2, we used an observed *trajectory* (one realization) $\{\tilde{x}_i, i = 1, 2, \ldots, n\}$ to estimate $\mathcal{A}$. $\tilde{\mathcal{A}}$ was then an *estimate*, based on the trajectory $\tilde{x}$. This was for presentational clarity, especially when writing the likelihood functions, but we shall immediately think of the trajectory as being random and consider that $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{B}}$ are random variable, *i.e.* estimators.

5.3   Joint estimation of $\mathcal{A}$ and α

In section 5.2, we assumed that α was known for the estimation of $\mathcal{A}$. Simultaneous estimation of $\mathcal{A}$ and α can be done using for example the Berndt Hall Hall Hausman (BHHH) algorithm or more advanced techniques [25], [93]. We will not explore these for two reasons. The first one is that they require to compute the derivative of the likelihood function (with exact transition probabilities) with respect to α, which has no closed-form expression due to the Bessel function in equation (5.44). The derivative with respect to $\mathcal{A}$ is very tedious but tractable. The second reason is that α can be estimated as explained previously without any knowledge on $\mathcal{A}$. A more natural and meaningful method to estimate both α and $\mathcal{A}$ emerges:

i use the ergodic estimator (equation (5.13)) to estimate α and get $\tilde{\alpha}_x$

ii estimate $\mathcal{A}$ using Euler's approximation or the exact transition probabilities, with α replaced by $\tilde{\alpha}_x$.

We set $\alpha = 1$ and $\mathcal{A} = 1$ Hz. As previously, we generate 1000 trajectories of a fixed duration T and estimate α and $\mathcal{A}$ as explained for each of the trajectories. We compute the estimation bias and standard deviation of α and $\mathcal{A}$ as a function of the duration T, for a duration between 10 and 1000 s. Figure 5.6 represents the results of this simulation. At known α, the estimation of $\mathcal{A}$ was satisfying even with 1 s long trajectories (section 5.2.4). However, when α must also be estimated, a longer trajectory is necessary as evidenced here. If the trajectory is too short, the error on α is large, and as a consequence so is that on
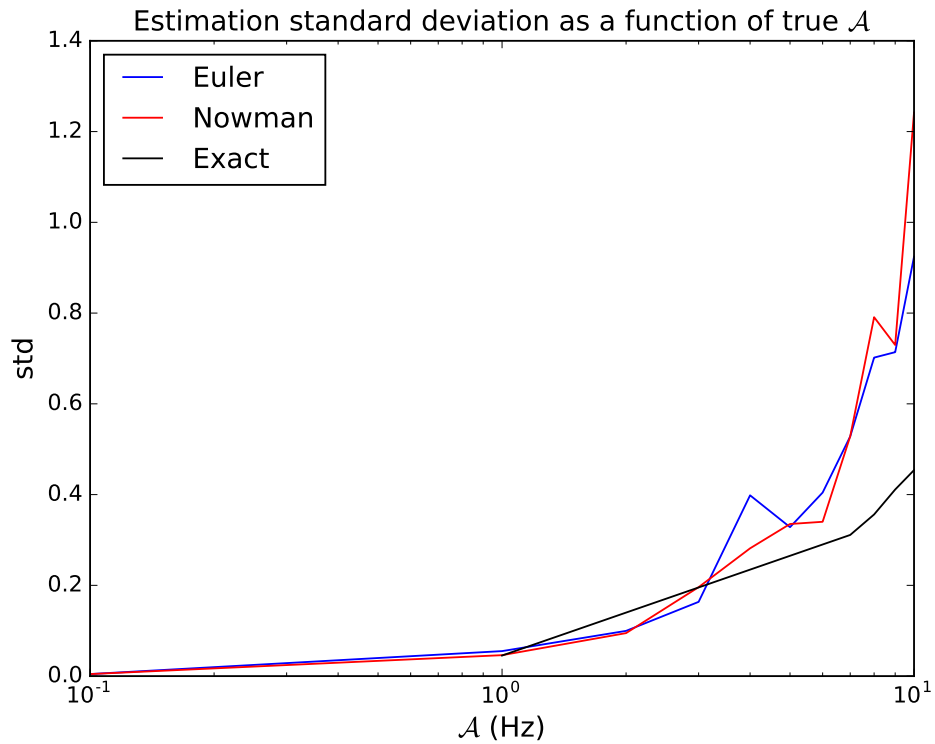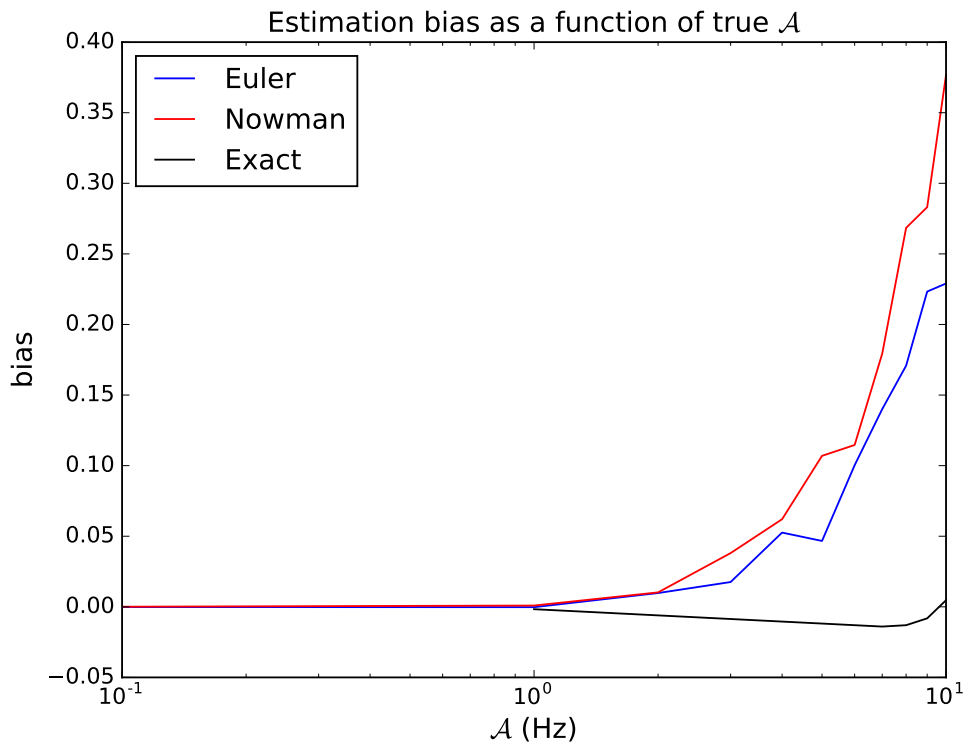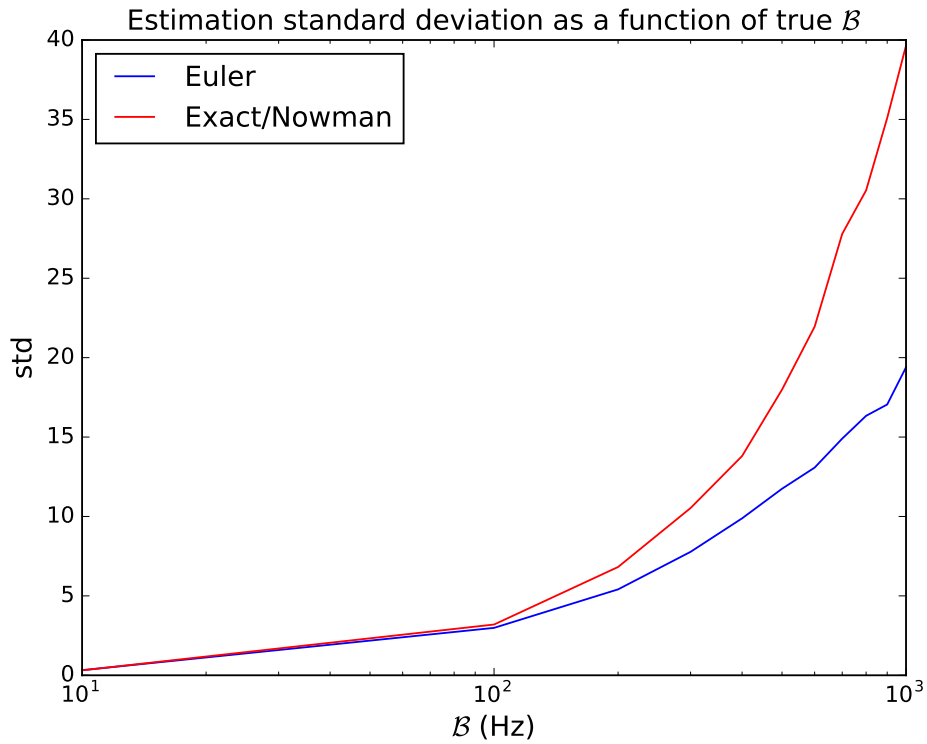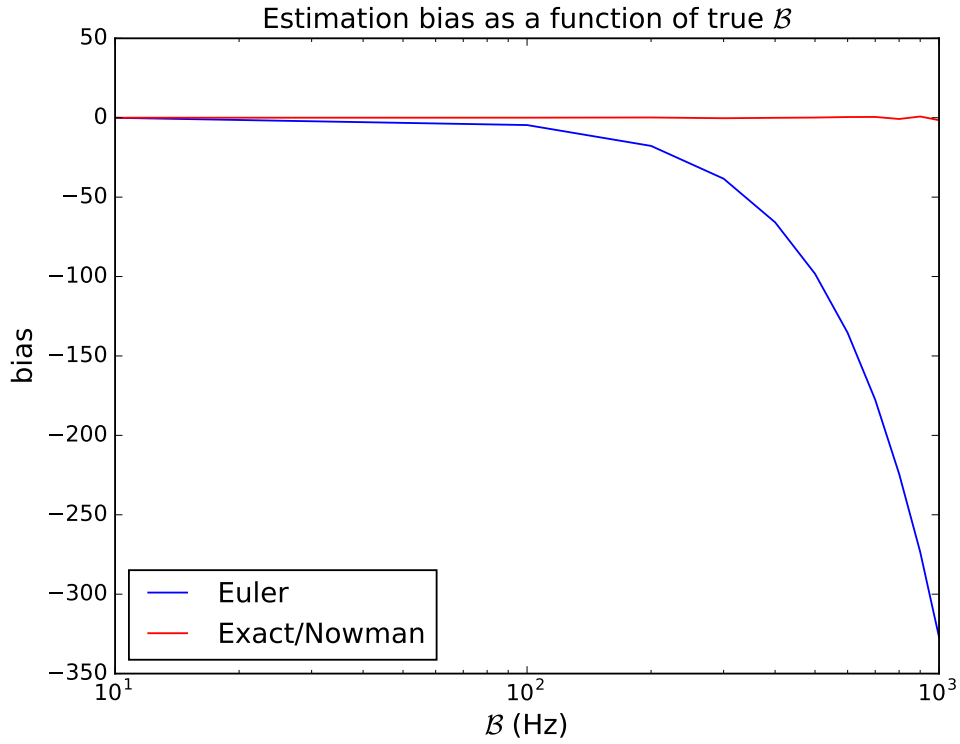
Figure 5.5: Estimation bias (up) and standard deviation (down) of $\mathcal{B}$ as a function of true $\mathcal{B}$ based on 1000 trajectories of duration 1 s each. 2 methods are compared: the Euler approximation and the Nowman approximation which is equal to the exact transition probability in this particular case.

$\mathcal{A}$. Surprisingly, the bias and standard deviation curves of $\mathcal{A}$ and α superimpose perfectly suggesting a correlation between their respective estimators.

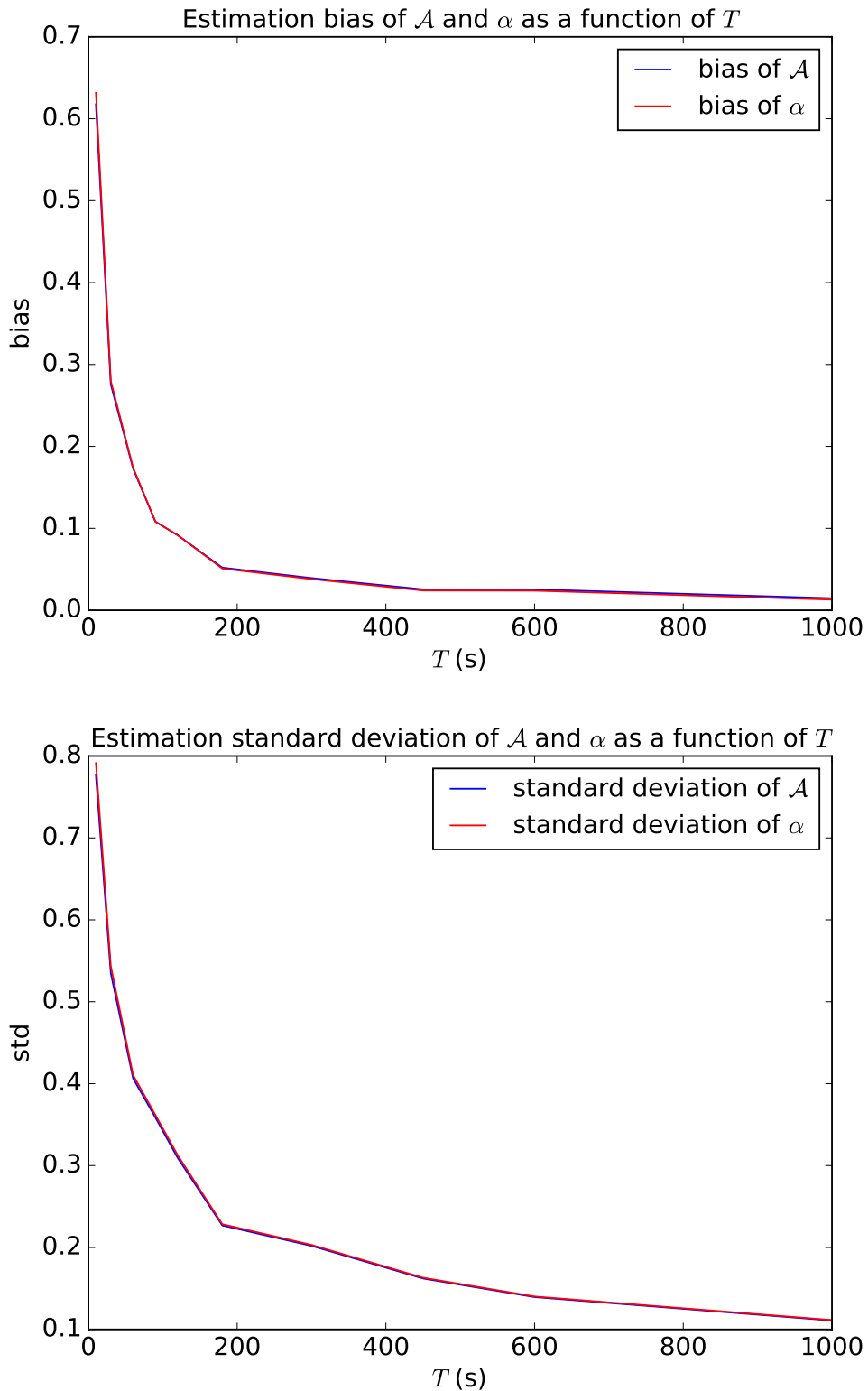
Figure 5.6: Estimation bias (up) and standard deviation (down) of $\mathcal{A}$ and α as a function of true the trajectory duration T, for $\mathcal{A} = 1$ Hz and $\alpha = 1$.

5.4 Discussion on the performance of the estimators

5.4.1 Errors on x_t from errors on the estimated $\mathcal{A}$

We saw in section 5.2, that using the exact transition probabilities in the maximum likelihood function is better than using Euler's approximation for at least three different reasons: it removes the bias, it reduces the variance (almost by a factor 2 for $\mathcal{A} = 10\,\mathrm{Hz}$, see figure 5.4), and we know *a priori* that it should be better. However, using the most complete model is not always necessary. We will show in this section that using Euler's approximation is sufficient to estimate $\mathcal{A}$. We use numerical simulations to convert an error made on the estimated parameter into an error made on the RCS x_t. We argue that the estimation is ultimately limited by the information carried in the trajectory that we use (sampling rate and duration).

Let us assume that an error $\delta\mathcal{A}$ is made on the estimation of $\mathcal{A}$, *i.e.* $\tilde{\mathcal{A}} = \mathcal{A} + \delta\mathcal{A}$. To convert it into errors on the RCS, we generate numerically trajectories of the RCS with the parameter $\mathcal{A}$ using Milstein's scheme. Its order of convergence is higher than Euler-Maruyama's but it follows the same principle: first we generate a series of brownian increments $\Delta W_{\hat{t}_k}, k = 1, 2, \ldots, 100n$, second we compute the corresponding series of the RCS $\hat{x}(\mathcal{A}) = x_{\hat{t}_k}(\mathcal{A}), k = 1, 2, \ldots, 100n$ starting from an arbitrary initial condition $x_{\hat{t}_0}$. Using the exact same brownian increments and initial condition, we can regenerate the trajectory with the estimated parameter and we obtain $\hat{x}(\mathcal{A} + \delta\mathcal{A}) = x_{\hat{t}_k}(\mathcal{A} + \delta\mathcal{A}), k = 1, 2, \ldots, 100n$. The trajectories $\hat{x}(\mathcal{A})$ and $\hat{x}(\mathcal{A} + \delta\mathcal{A})$ are expected to be identical for small $\delta\mathcal{A}$.

Figure 5.7 represents the results of this method applied to the RCS in the case where α is known and equal to 1. We generate 1000 trajectories of the RCS, $\hat{x}^{(i)}(\mathcal{A})$, $i = 1, 2, \ldots, 1000$, $1\,\mathrm{s}$ long each and save the brownian increments used to generate them. $\mathcal{A}$ is estimated from each trajectory, which gives $\tilde{\mathcal{A}}_i$, $i = 1, 2, \ldots, 1000$. The estimation is carried out using the exact transition probabilities (unbiased estimation) and Euler's approximation. For the latter, we remove the bias by indirect inference [110]: our previous simulations already gave us the bias for $\mathcal{A} = 1\,\mathrm{Hz}$. The trajectories are regenerated with the estimated parameters, which gives $\hat{x}^{(i)}(\tilde{\mathcal{A}}_i)$, $i = 1, 2, \ldots, 1000$. Finally, we compute the mean standard deviation between the trajectories:

$$\sigma_{est} = \frac{1}{1000} \sum_{i=1}^{1000} \sigma_i \tag{5.45}$$

with

$$\sigma_i^2 = \frac{1}{100n - 1} \sum_{k=1}^{100n} \left(x_{\hat{t}_k}^{(i)}(\mathcal{A}) - x_{\hat{t}_k}^{(i)}(\tilde{\mathcal{A}}_i) \right)^2. \tag{5.46}$$

The lower part of Figure 5.7 shows the standard deviation between the original and regenerated trajectories, with a true parameter $\mathcal{A} = 1\,\mathrm{Hz}$. We plot only the first 100 values to clearly show the correlation. They are pretty much the same for the trajectories regenerated with Euler's estimation of the parameter $\mathcal{A}$, and those regenerated with the estimation of the parameter using the exact transition probabilities (see section 5.2). In these simulations, the mean standard deviation is $6.9.10^{-3}$ for Euler's approximation and $6.77.10^{-3}$ for the estimation based on the exact transition probabilities. The standard deviation of the difference between the regenerated trajectory using Euler's estimation and the estimation based on exact transition probabilities (black curve) is much smaller, with an average of

$9.2.10^{-4}$. This can be summed up by the upper part of figure 5.7: for both estimations, the regenerated trajectories are much closer together than they are to the original trajectory.

Figure 5.8 confirms this: it shows that the estimated parameters are highly correlated when plotted as a function of trajectory number. It follows immediately that the regenerated trajectories will be very close, as well as all our comments about figure 5.7. The estimations have been debiased for Euler's estimation by indirect inference, *i.e.* by an established bijection between the estimated value and the bias, based on the results in figure 5.4 (top). However, this was not necessary due to the very low bias at $\mathcal{A} = 1$ Hz. Our interpretation is that the error on the estimation of $\mathcal{A}$ in our configuration (1 s long time series sampled at 1 ms) is mainly due to the statistical peculiarity of the trajectory, not the choice between Euler's approximation or the exact transition probabilities or maybe any other estimation method. The time series is one short sampled chunk of one time-continuous and of infinite duration realization of the random process, and therefore carries only limited information that may be almost completely used in a method as simple as Euler's approximation.

Out of curiosity, we compared the standard deviations of Figure 5.7 to the typical measurement standard deviation of a satellite radar sensor. The RCS is normally in units of a surface, generally m². Taking the decimal logarithm and multiplying by 10 gives the RCS expressed in dBm². Antony et al. [14] have shown that the post-calibration radiometric accuracy (or measurement standard deviation) of the TerraSAR-X and TanDEM-X satellite systems is of 0.25 dBm² for a target 43.5 dBm². By definition, x_t is the ratio of the observed sea-surface RCS (expressed as surfaces) to its mean value, we can show that the measurement standard deviation of x_t corresponding to 0.25 dBm² for a target of 43.5 dBm² is $4.73.10^{-5}$, much smaller than $6.9.10^{-3}$ and $6.77.10^{-3}$. Our estimation, regardless of the used method, is not within radiometric accuracy when converted into sea-clutter units.

5.4.2 RMSE after bias correction

In section 5.2, we express ML estimators and numerically compute their bias and standard deviations to measure how well they perform. In the previous section, we focused on $\mathcal{A}$ and noticed that Euler's approximation and the exact transition probabilities yield almost identical estimations on a trajectory to trajectory basis (figure 5.8). We set $\mathcal{A} = 1$ Hz, *i.e.* the reference value quoted for $\mathcal{A}$. At such a value, the bias is almost zero for Euler's approximation, and its standard deviation almost identical to that of the exact transition probabilities estimator (see figure 5.4), so it is not surprising that both estimator perform equally well.

However, it is not as easy to draw conclusions for $\mathcal{B}$. Figure 5.5 shows that for example for $\mathcal{B} = 1000$ Hz, the bias of Euler's estimator is -350 Hz while that of the exact transition probabilities estimator is 0 Hz. For the standard deviation, Euler's is half that of the exact transition probabilities estimator! How can we draw a conclusion from this observation? We think that the relevant measure of performance is the root mean square error after debiasing.

First, let us explicit the debiasing procedure. We explicit it using the notation $\mathcal{B}$ but it is general for any parameter. So let $\mathcal{B}$ be the real value of the parameter (deterministic). Let $\tilde{\mathcal{B}}(\mathcal{B})$ be an estimator of $\mathcal{B}$ (random variable). For example, it can be the ML estimator with Euler's approximation, for a 1 s long trajectory of $\gamma_t^{(R)}$ sampled at 0.001 s. The expectation of the estimator is:

$$\mathbb{E}[\tilde{\mathcal{B}}(\mathcal{B})] = \mathcal{B} + b(\mathcal{B}), \tag{5.47}$$
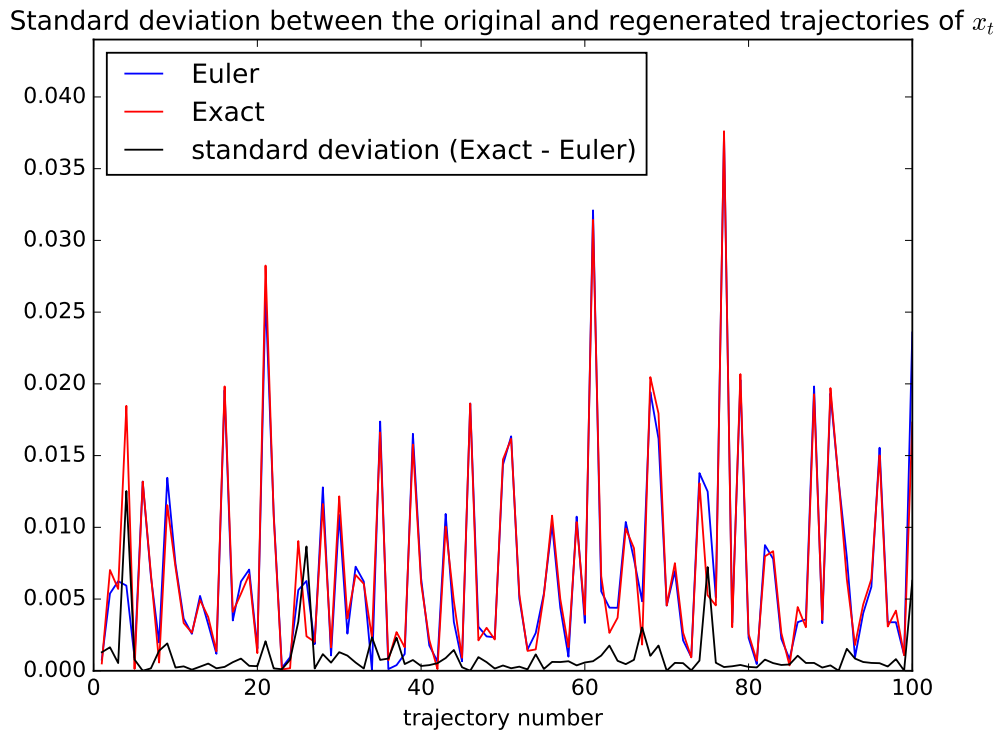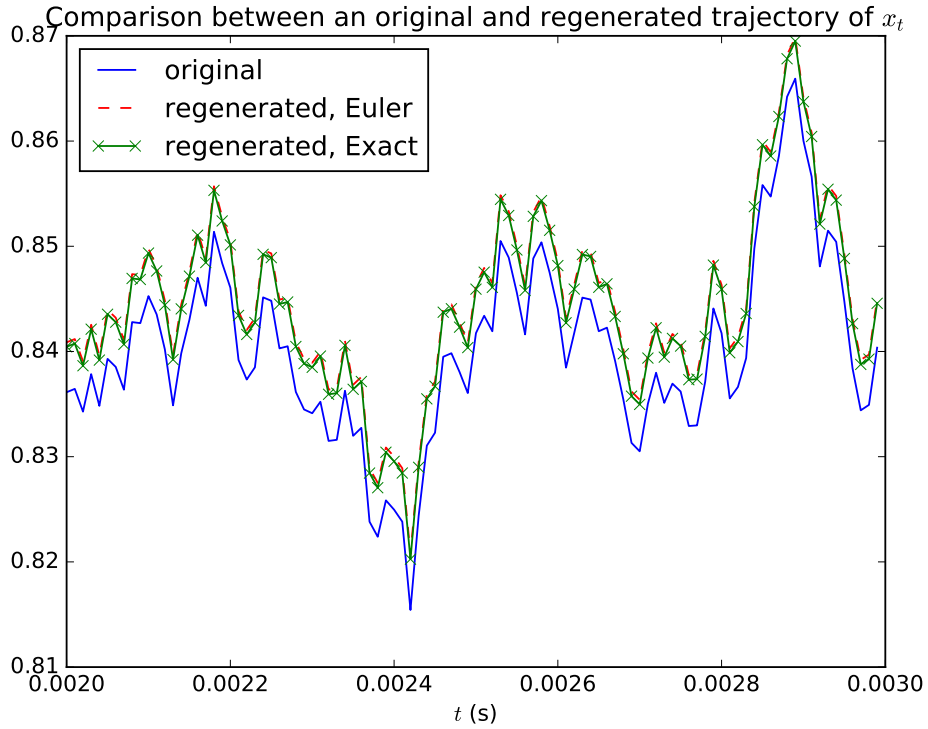
Figure 5.7: Up: example of an original trajectory generated with $\mathcal{A} = 1\,\mathrm{Hz}$ represented along with its regenerated counterparts after estimation of $\mathcal{A}$ using Euler's approximation and the exact transition probabilities. Down: standard deviation of the difference between the original and regenerated trajectory as a function of the trajectory number.
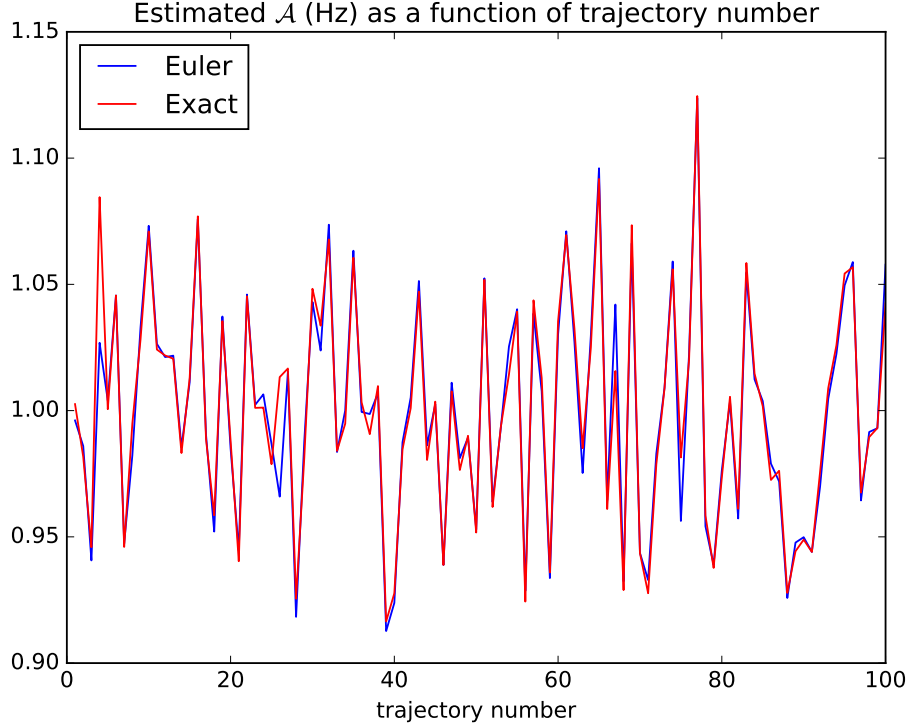
Figure 5.8: Comparison between the estimated $\mathcal{A}$ Euler's approximation and the exact transition probabilities as a function of trajectory number.

where b is the bias function. The standard deviation of the estimator is:

$$\sigma(\mathcal{B}) = \mathbb{E}\left[(\tilde{\mathcal{B}}(\mathcal{B}) - \mathbb{E}[\tilde{\mathcal{B}}(\mathcal{B})])^2\right]^{1/2} = \mathbb{E}\left[(\tilde{\mathcal{B}}(\mathcal{B}) - \mathcal{B} - b(\mathcal{B}))^2\right]^{1/2}. \tag{5.48}$$

The bias $b(\mathcal{B})$ and standard deviation $\sigma(\mathcal{B})$ as defined in equations (5.47) and (5.48) are what we represented in figures 5.4 for $\mathcal{A}$ and 5.5 for $\mathcal{B}$ (more precisely, we represent estimations of the estimation bias and standard deviation). The RMSE for the biased estimator $\tilde{\mathcal{B}}(\mathcal{B})$ is:

$$\mathrm{rmse}(\mathcal{B}) = \mathbb{E}\left[\left(\tilde{\mathcal{B}}(\mathcal{B}) - \mathcal{B}\right)^2\right]^{1/2}. \tag{5.49}$$

Since $\mathcal{B} = \mathbb{E}[\tilde{\mathcal{B}}(\mathcal{B})] - b(\mathcal{B})$, the squared RMSE can be decomposed:

$$
\begin{aligned}
\mathrm{rmse}(\mathcal{B})^2 &= \mathbb{E}\left[\left(\tilde{\mathcal{B}}(\mathcal{B}) - \mathcal{B}\right)^2\right] = \mathbb{E}\left[\left(\tilde{\mathcal{B}}(\mathcal{B}) - \mathbb{E}[\tilde{\mathcal{B}}(\mathcal{B})] + b(\mathcal{B})\right)^2\right] \\
\Leftrightarrow \mathrm{rmse}(\mathcal{B})^2 &= \mathbb{E}\left[(\tilde{\mathcal{B}}(\mathcal{B}) - \mathbb{E}[\tilde{\mathcal{B}}(\mathcal{B})])^2\right] + b(\mathcal{B})^2 + 2b(\mathcal{B})\mathbb{E}\left[(\tilde{\mathcal{B}}(\mathcal{B}) - \mathbb{E}[\tilde{\mathcal{B}}(\mathcal{B})])\right] \\
\Leftrightarrow \mathrm{rmse}(\mathcal{B})^2 &= \sigma(\mathcal{B})^2 + b(\mathcal{B})^2. \tag{5.50}
\end{aligned}
$$

The RMSE is therefore a measure of distance between the estimator and the true parameter which takes into account both the bias and standard deviation of the estimator. If we get back to Euler's approximation and the exact transition probabilities estimators, their RMSE squared are (equation (5.50)) respectively $20^2 + (-350)^2 = 122900$ Hz^2 and $40^2 + 0^2 = 1600$ Hz^2. We see a clear advantage for the exact transition probabilities, which is of course due

135

to the large bias of the other estimator. This direct comparison is not fair, because the bias can be partially corrected if one knows the bias function $b(\mathcal{B})$. To compute the estimated bias $\tilde{b}(\mathcal{B})$, we first inverse:

$$\tilde{\mathcal{B}}(\mathcal{B}) = \hat{\mathcal{B}}(\mathcal{B}) + b(\hat{\mathcal{B}}(\mathcal{B})) \tag{5.51}$$

to obtain $\hat{\mathcal{B}}(\mathcal{B})$, which is an estimation of the true $\mathcal{B}$ that would generate the observed value $\tilde{\mathcal{B}}(\mathcal{B})$ if it was sure that $\tilde{\mathcal{B}}(\mathcal{B}) = \mathbb{E}\left[\tilde{\mathcal{B}}(\mathcal{B})\right]$. The estimated bias is then:

$$\tilde{b}(\mathcal{B}) = b(\hat{\mathcal{B}}). \tag{5.52}$$

The bias corrected estimator is simply:

$$\tilde{\mathcal{B}}(\mathcal{B}) - \tilde{b}(\mathcal{B}) = \hat{\mathcal{B}}(\mathcal{B}). \tag{5.53}$$

If $\tilde{\mathcal{B}}(\mathcal{B}) = \mathbb{E}\left[\tilde{\mathcal{B}}(\mathcal{B})\right]$, then equation (5.51) is:

$$\mathbb{E}\left[\tilde{\mathcal{B}}(\mathcal{B})\right] = \hat{\mathcal{B}}(\mathcal{B}) + b(\hat{\mathcal{B}}(\mathcal{B})), \tag{5.54}$$

the solution of which is $\hat{B}(\mathcal{B}) = \mathcal{B}$. In that case, we obtain the exact value of $\mathcal{B}$. However, since $\tilde{\mathcal{B}}(\mathcal{B})$ takes values around its expectation $\mathbb{E}\left[\tilde{\mathcal{B}}(\mathcal{B})\right]$ according to its standard deviation $\sigma(\mathcal{B})$, the corrected bias is most likely never exactly $b(\mathcal{B})$. Therefore, the standard deviation of the estimator induces errors in the estimated bias, such that the bias corrected estimator has variance greater than the unbiased estimator.

This is a crucial remark since if perfect bias correction was possible, then only the standard deviation of the biased estimator would matter. In that case, we would deduce that Euler's approximation ML estimator is the best for $\mathcal{B}$ from figure 5.5. Instead, it is important to be aware that there is a bias/standard deviation interaction in the bias correction, which leads to an unbiased estimator $\hat{B}(\mathcal{B})$ of unknown standard deviation. Since the bias correction that leads to $\hat{B}(\mathcal{B})$ is the best we can do, we can now define the rmse of $\hat{B}(\mathcal{B})$:

$$\text{rmse}(\hat{B}(\mathcal{B})) = \mathbb{E}\left[\left(\hat{B}(\mathcal{B}) - \mathcal{B}\right)^2\right]^{1/2}. \tag{5.55}$$

Equation (5.55) is the correct measure for the performance of the estimator. We have done new numerical simulations to compute the RMSE of Euler's approximation and exact transition probabilities ML estimators for both $\mathcal{A}$ and $\mathcal{B}$. For $\mathcal{A}$, the range of values goes from 0.1 Hz to 10 Hz with a step of 0.9 Hz from 0.1 to 1 Hz and then steps of 1 Hz from 1 to 10 Hz. For $\mathcal{B}$, the range of values goes from 10 Hz to 1000 Hz with a step of 90 Hz from 10 to 100 Hz and then steps of 100 Hz from 100 to 1000 Hz. For each value, 1000 trajectories are simulated (of x_t for $\mathcal{A}$ and of $\gamma_t^{(R)}$ for $\mathcal{B}$) and the parameters are estimated. The bias is corrected according to the above procedure and the RMSE is then computed.

Figure 5.9 represents the results for $\mathcal{B}$. It is striking to see that the RMSE for the two estimators are almost identical. The RMSE for the exact transition probabilities estimator is the same as its standard deviation since it is unbiased. However, the RMSE of the Euler's approximation estimator is greater than its standard deviation due to the imperfectness of the bias correction. The increase is exactly such that it reaches the standard deviation of the other estimator. We can now conclude that Euler's approximation and the exact transition probabilities are equally good for estimating $\mathcal{B}$.
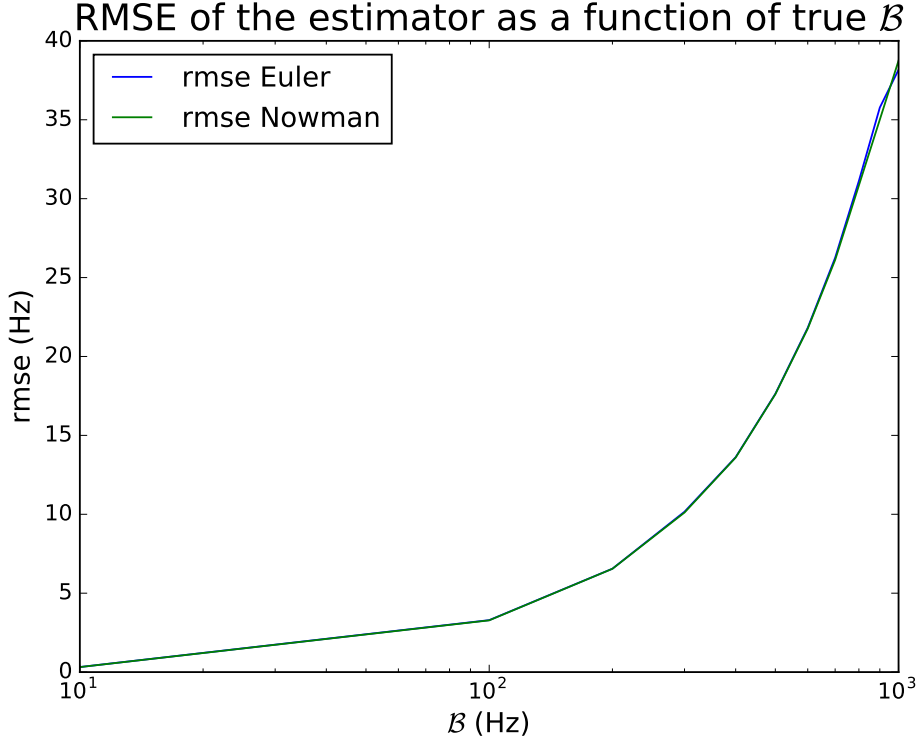
Figure 5.9: Root mean square error of the estimators for $\mathcal{B}$ as a function of true $\mathcal{B}$.

5.5   Spectrum-based estimation of $\mathcal{B}$

It has been suggested by an anonymous reviewer of our publication [121] that the bandwidth of z_t may be used to estimate $\mathcal{B}$. The intuition is that the bandwith of the intensity is dictated by that of the speckle squared $|\gamma_t|^2$ since the speckle has dynamics much faster than those of the RCS x_t. As a consequence, the high-frequency content would depend mostly on γ_t and therefore of $\mathcal{B}$. In this section, we follow this trail to see if it indeed provides a new way to estimate $\mathcal{B}$. Field in [48] chapter 8 shows that:

$$\langle z_\tau z_0 \rangle = \left(1 + \frac{1}{\alpha} e^{-\mathcal{A}\tau} \right) \left(1 + e^{-\mathcal{B}\tau} \right). \tag{5.56}$$

In the present context, $\langle . \rangle$ is a classical notation for the mathematical expectation $\mathbb{E}$. If both positive and negative τ are considered and if we develop equation (5.56), we get:

$$\langle z_\tau z_0 \rangle = 1 + e^{-\mathcal{B}|\tau|} + \frac{1}{\alpha} e^{-\mathcal{A}|\tau|} + \frac{1}{\alpha} e^{-(\mathcal{A}+\mathcal{B})|\tau|}. \tag{5.57}$$

By Wiener-Khinchin theorem (see chapter 2 section 2.6), the power-spectral density (PSD) of z is:

$$S_z(\omega) = \frac{1}{2\pi} \int_{\mathbb{R}} \langle z_\tau z_0 \rangle e^{-i\omega\tau} d\tau. \tag{5.58}$$

Since it holds that for all a, $\int_{\mathbb{R}} e^{-a|\tau|} e^{-i\omega\tau} d\tau = \frac{2a}{a^2+\omega^2}$, we have:

$$S_z(\omega) = \frac{1}{2\pi} \delta_0 + \frac{\mathcal{B}}{\pi \left(\mathcal{B}^2 + \omega^2 \right)} + \frac{1}{\alpha} \frac{\mathcal{A}}{\pi \left(\mathcal{A}^2 + \omega^2 \right)} + \frac{1}{\alpha} \frac{\mathcal{A} + \mathcal{B}}{\pi \left((\mathcal{A} + \mathcal{B})^2 + \omega^2 \right)}, \tag{5.59}$$

where δ_0 is the Dirac distribution. Ideally, we would like to observe that $S_z(\omega)$ can be approximated by a function where only the parameter $\mathcal{B}$ appears. If we compare $\frac{\mathcal{B}}{(\mathcal{B}^2+\omega^2)}$ and $\frac{\mathcal{A}}{(\mathcal{A}^2+\omega^2)}$:

$$\frac{\mathcal{A}}{(\mathcal{A}^2+\omega^2)} \bigg/ \frac{\mathcal{B}}{(\mathcal{B}^2+\omega^2)} = \frac{\mathcal{A}}{\mathcal{B}}\left(\frac{\mathcal{B}^2+\omega^2}{\mathcal{A}^2+\omega^2}\right) \xrightarrow{\omega\to+\infty} \frac{\mathcal{A}}{\mathcal{B}}. \tag{5.60}$$

Though it holds that $\frac{\mathcal{A}}{\mathcal{B}}$ is small, it is not equal to zero. More annoyingly, we may replace $\frac{1}{\alpha}\frac{\mathcal{A}+\mathcal{B}}{\pi((\mathcal{A}+\mathcal{B})^2+\omega^2)}$ by $\frac{1}{\alpha}\frac{\mathcal{B}}{(\mathcal{B}^2+\omega^2)}$, but a priori α is unknown as well as $\mathcal{A}$. Assuming for example that $\alpha = 1$, we propose the following approximation:

$$S_z(\omega) \approx \frac{1}{2\pi}\delta_0 + \frac{2\mathcal{B}}{\pi\left(\mathcal{B}^2+\omega^2\right)}. \tag{5.61}$$

Figure 5.10 represents the exact PSD (equation (5.59)) and the approximate PSD (equation (5.61) for three of $\mathcal{B}$: 50, 100 and 200 Hz. $\mathcal{A}$ and α are set to one. They are represented as functions of $f = \omega/2\pi$. The difference between the approximate and exact PSD is large compared to the difference for different $\mathcal{B}$. For example, the exact PSD for $\mathcal{B} = 50$ Hz is closer to the approximate PSD for 100 Hz than 50 Hz. For comparison, the PSD of a simulated trajectory of z_t of duration 300 s and timestep $\Delta t = 10^{-3}$ s is also computed with the *signal.welch* function of Python. Based on these numerical results, it seems that there is unfortunately no convenient way to estimate $\mathcal{B}$ from the PSD of a trajectory without knowing $\mathcal{A}$ and α.
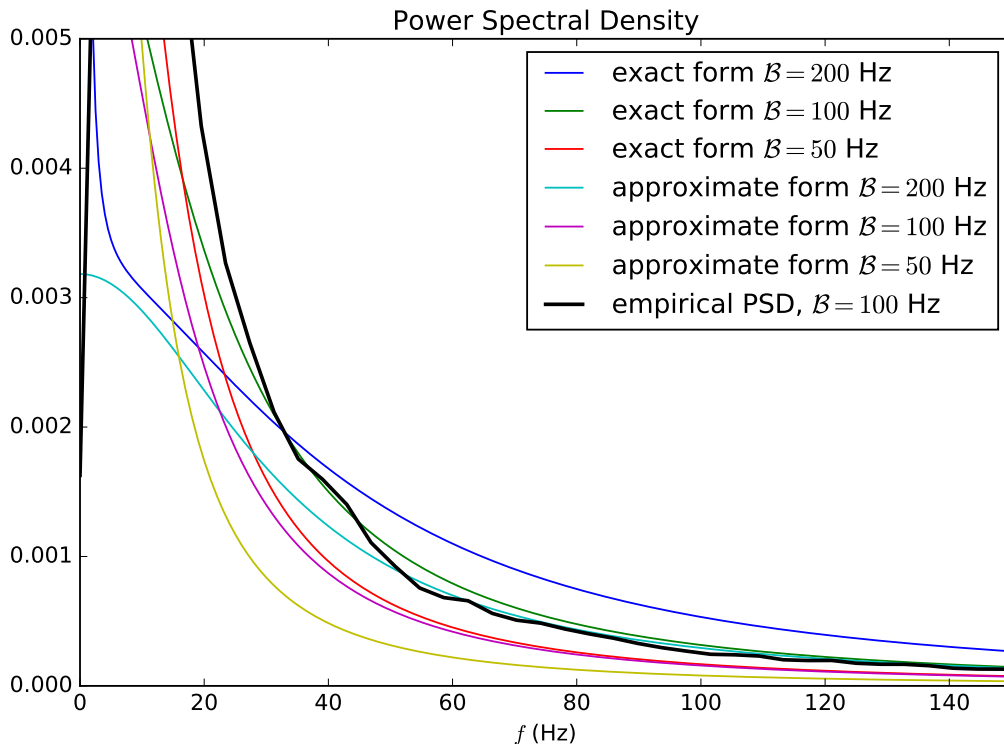
Figure 5.10: Comparison between exact, approximate and empirical PSDs as a function of the frequency f. $\mathcal{A}$ and α are set to one. The duration of the trajectory for the empirical PSD is 300 s and the timestep 10^{-3} s.

5.6 Conclusion

In this chapter, we took advantage of our previous results to address the issue of estimating the three parameters of Field's model (equation 5.3): $\mathcal{A}$, $\mathcal{B}$ and α. α is estimated using the ergodic property of the sea clutter (section 5.1) either from x_t (equation (5.13)), or from z_t (equation (5.21)). The estimator from x_t, $\tilde{\alpha}_x$, is better if x_t is observed. In practice, only z_t is observed so the estimator from z_t, $\tilde{\alpha}_z$, is very useful (see chapter 7). $\mathcal{A}$ and $\mathcal{B}$ are estimated by maximum likelihood (section 5.2). We address the issue of estimating x_t from observable quantities in chapter 7, section 7.1. We also show in chapter 7 that if x_t has to be estimated, the estimation of $\mathcal{A}$ is more tricky.

We compared three methods for the estimation of $\mathcal{A}$ and $\mathcal{B}$ (section 5.2). Euler's approximation assumes a constant drift and volatility over small intervals, it leads to an explicit formula for the estimator. Nowman's approximation resolves the drift but assumes constant volatility over small intervals, and leads also to an explicit formula. In contrary, using the exact transition probabilities makes no assumption but does not lead to an explicit formula. It requires numerical minimization. It was observed in section 5.2 that Nowman's method has the same performance as Euler's, so this method was discarded straight away for estimating $\mathcal{A}$ in the rest of the chapter. For estimating $\mathcal{B}$, we remind that it is strictly the same as using the exact transition probabilities due to the fact that γ_t has constant volatility. Finally, we have also proposed a convenient joint estimation of $\mathcal{A}$ and α (section 5.3).

Throughout the chapter, we used numerical simulations to compare the respective performance of our estimators, in connection with the specificities of the application. In our application, the specificities were the numerical values of the timestep and the duration of the time series, as well as realistic values for the parameters. We quoted that $\mathcal{A}$ is around 1 Hz, $\mathcal{B}$ is around 100 Hz, and α is around 1 (sections 5.2 and 5.1.2). We showed that even though using the exact transition probabilities gives the best results, Euler's approximation is sufficient provided that the estimator is debiased by indirect inference. In practice, there is indeed a strong correlation between these two estimations as a function of the trajectory (figure 5.8) which suggests that it is the trajectory itself which limits the performance of the estimation due to its limited information about the parameters.

Chapter 6

Estimation of target parameters in sea clutter

Detecting the presence of a coherent scatterer (target) in a time-evolving random medium by remote sensing can be challenging, in particular if the field scattered by the dynamic random medium (clutter), which is unpredictable in nature, has a mean power of the same order of magnitude as the field scattered by the target. It is also a significant concern with many real life applications. One striking example is that of a small boat on the sea surface. Detecting the boat (coherent scatterer) into a strong sea clutter (random medium) by radar is still challenging due to the dynamics of the sea surface [34]. It is also of high value for maritime surveillance concerns. A slightly different example is the detection of a boat (or other floating objects) from underwater by sonar, in the context of submarines or underwater autonomous vehicles surfacing [80]. In this case, the clutter comes from the reflection of the sonar waves with the sea surface, bubbles generated by breaking waves, or even algae.

Target detection based on the random walk model can be found in [141] chapter 6. As explained, in section 1.3, this model cannot describe precisely the dynamics of the clutter, contrary to Field's model [48]. It is therefore natural to think that Field's model may provide new insights for target detection.

We remind that in absence of a target, Field's model describes the (normalized) complex reflectivity of a random medium (*e.g.* the sea surface) as the product:

$$\Psi_t = x_t^{1/2}\gamma_t \tag{6.1}$$

of the square root of the radar cross section (RCS) x_t and the speckle γ_t. If there is a target, the reflectivity becomes:

$$\Psi_t^{(target)} = x_t^{1/2}\gamma_t + \Psi_{c,t} \tag{6.2}$$

where $\Psi_{c,t}$ is the time-dependent reflectivity of the target. We consider for this chapter that the target is present during the whole duration of the trajectory and focus on two models for its reflectivity, Homodyned K (HK, constant target) and Generalized K (GK, constant multiplied by the RCS), see [48] chapter 9 and [141] p 112. For HK and GK scattering, the target is parameterized by two parameters.

The HK and GK models have not been initially developed for target detection, but for *weak scattering*. Weak scattering happens when the scatterers in the random walk do not have a phase with uniform distribution over $[0, 2\pi[$ but have it distributed close to some mean value [71]. It is equivalent to adding a bias in the random walk and can be relevant

to the general case of diffusion by a phase screen (see also JakemanRidley2006). Therefore, the work presented here is applicable also in very different contexts, for example to estimate weak scattering of laser by a layer of turbulent atmosphere [20].

In this chapter, we derive the stochastic differential equations (SDE) for $\Psi_t^{(target)}$ and estimate the target parameters by maximum likelihood (ML) under the assumption that the clutter parameters $\mathcal{A}$, $\mathcal{B}$ and α are known and that x_t is observed. In section 6.1, we use Field's model and Itô calculus to derive the SDE of the clutter plus target for HK and GK scattering. In section 6.2, we derive analytical approximations for the transition probabilities in HK and GK scattering using Euler-Maruyama scheme. We then show that maximum likelihood estimation of the target is possible using these transition probabilities and leads to explicit formula. In section 6.3, we assess the performance of the maximum likelihood estimator using numerical simulations of the SDE we derived in section 6.1 as compared to a simple ergodicity-based estimator. We assign numerical values to the constants of the model based on the sea surface application. Section 6.4 is a discussion about the limits and advantages of our approach, as well as its implications in terms of decision theory. Finally, we conclude in section 6.5.

Because Field's model describes the complex reflectivity of an arbitrary random medium, the results of this chapter are not limited to target estimation in sea clutter: they may concern target parameter estimation in any random medium.

In appendix A.2, we derive volatility-based estimators for Ψ_c. In HK scattering, volatility-based estimation has lesser performance (compared to ML), and in GK scattering the performance is similar. As for $\mathcal{A}$ and $\mathcal{B}$ in the previous chapter, the volatility-based estimators are eventually inoperative. Despite the non-observability of x_t which degrades the sampling frequency (chapter 7), its poor performance for HK scattering forces us to dismiss the volatility-based estimators. However, we again recommend to read the appendix after this chapter due to its striking simplicity compared to the derivations of the ML estimators.

6.1   SDE of the clutter plus target

In [48], chapter 9, the SDE of the clutter plus target are expressed in polar coordinates, which yields heavy expressions. In this section, we derive the SDE of the clutter plus target (coherent scatterer) in cartesian coordinates (real and imaginary parts) since it is a more convenient coordinate system for target estimation. Two models for the target reflectivity are considered: Homodyned K scattering and Generalized K scattering (see [48] p 70-71 and see sections 6.1.2 and 6.1.3). It will be evidenced in equations (6.10) and (6.14) that HK and GK scattering encompass the cases of Rice, Rayleigh and K distribution scattering. Rice scattering is recovered when there is a target and the RCS x_t is constant, Rayleigh scattering is recovered when there is no target and the RCS is constant, and K distribution scattering is recovered when there is no target but still a varying RCS.

We first remind that Field's model for the complex reflectivity $\Psi_t = x_t^{1/2}\left(\gamma_t^{(R)} + i\gamma_t^{(I)}\right)$ (clutter) can be expressed as the set of three SDE:

$$
\begin{cases}
\mathrm{d}x_t = \mathcal{A}(1 - x_t)\mathrm{d}t + \left(2\frac{\mathcal{A}}{\alpha}x_t\right)^{\frac{1}{2}}\mathrm{d}W_t^{(x)} \\
\mathrm{d}\gamma_t^{(R)} = -\frac{1}{2}\mathcal{B}\gamma_t^{(R)}\mathrm{d}t + \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\mathrm{d}W_t^{(R)} \\
\mathrm{d}\gamma_t^{(I)} = -\frac{1}{2}\mathcal{B}\gamma_t^{(I)}\mathrm{d}t + \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\mathrm{d}W_t^{(I)}.
\end{cases}
\tag{6.3}
$$

The three constant parameters $\mathcal{A}$, $\mathcal{B}$ and α have been estimated in chapter 5.

6.1.1 SDE of the real and imaginary parts of the clutter

As a preliminary step, we derive the SDE of the in-phase (real part) component, denoted $R_t^{(cl)}$, and quadrature phase (imaginary part) component, denoted $I_t^{(cl)}$, of the clutter only. We have $R_t^{(cl)} = x_t^{1/2} \gamma_t^{(R)}$ and $I_t^{(cl)} = x_t^{1/2} \gamma_t^{(I)}$. Since from equation (6.3) $\gamma_t^{(R)}$ and $\gamma_t^{(I)}$ follow the same SDE with different brownian motions, we just need to derive the SDE for $R_t^{(cl)}$ for example, and that of $I_t^{(cl)}$ will follow immediately.

Let $s_t = x_t^{1/2}$. Using Itô's formula for $f(x_t)$ with $f(x) = x^{1/2}$ (proposition 2.6), we have:

$$\mathrm{d}s_t = \frac{\partial f}{\partial x}(x_t)\mathrm{d}x_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x_t)\mathrm{d}\langle x\rangle_t. \tag{6.4}$$

$\mathrm{d}\langle x\rangle_t$ is the stochastic differential of the quadratic variation of x_t (see section 2.4). We get:

$$\begin{aligned}
\mathrm{d}s_t &= \frac{1}{2x_t^{1/2}}\left(\mathcal{A}(1-x_t)\mathrm{d}t + \left(2\frac{\mathcal{A}}{\alpha}x_t\right)^{\frac{1}{2}}\mathrm{d}W_t^{(x)}\right) - \frac{1}{8x_t^{3/2}}\frac{2\mathcal{A}}{\alpha}x_t\mathrm{d}t \\
&= \frac{\mathcal{A}}{2}\left(\frac{1}{s_t} - s_t\right)\mathrm{d}t + \left(\frac{\mathcal{A}}{2\alpha}\right)^{1/2}\mathrm{d}W_t^{(x)} - \frac{\mathcal{A}}{4\alpha s_t}\mathrm{d}t.
\end{aligned} \tag{6.5}$$

Thus we obtain:

$$\mathrm{d}s_t = \frac{\mathcal{A}}{2}\left(\frac{1}{s_t}\left(1 - \frac{1}{2\alpha}\right) - s_t\right)\mathrm{d}t + \left(\frac{\mathcal{A}}{2\alpha}\right)^{1/2}\mathrm{d}W_t^{(x)}. \tag{6.6}$$

From Itô's product law (proposition 2.5), we get:

$$\begin{aligned}
\mathrm{d}R_t^{(cl)} &= s_t\mathrm{d}\gamma_t^{(R)} + \gamma_t^{(R)}\mathrm{d}s_t + \mathrm{d}\langle s, \gamma^{(R)}\rangle_t \\
\Leftrightarrow \mathrm{d}R_t^{(cl)} &= s_t\mathrm{d}\gamma_t^{(R)} + \gamma_t^{(R)}\mathrm{d}s_t,
\end{aligned} \tag{6.7}$$

where we have used that $\langle s, \gamma^{(R)}\rangle_t$, the quadratic variation at time t of the processes s_t and $\gamma_t^{(R)}$, is zero by independence of x_t and $\gamma_t^{(R)}$. Using equation (6.7), the second equation in (6.3), and equation (6.6), we get:

$$\begin{aligned}
\mathrm{d}R_t^{(cl)} = {}& -\frac{1}{2}\mathcal{B}\gamma_t^{(R)}s_t\mathrm{d}t + \frac{s_t}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\mathrm{d}W_t^{(R)} \\
& + \frac{\mathcal{A}}{2}\left(\frac{1}{s_t}\left(1 - \frac{1}{2\alpha}\right) - s_t\right)\gamma_t^{(R)}\mathrm{d}t + \gamma_t^{(R)}\left(\frac{\mathcal{A}}{2\alpha}\right)^{1/2}\mathrm{d}W_t^{(x)}
\end{aligned}$$

which gives after factorization:

$$\begin{aligned}
\mathrm{d}R_t^{(cl)} = {}& \left(-\frac{\mathcal{A}+\mathcal{B}}{2}s_t\gamma_t^{(R)} + \frac{\mathcal{A}\gamma_t^{(R)}}{2s_t}\left(1 - \frac{1}{2\alpha}\right)\right)\mathrm{d}t \\
& + \left[\gamma_t^{(R)}\left(\frac{\mathcal{A}}{2\alpha}\right)^{1/2} \quad s_t\left(\frac{\mathcal{B}}{2}\right)^{1/2}\right]\begin{bmatrix}\mathrm{d}W_t^{(x)} \\ \mathrm{d}W_t^{(R)}\end{bmatrix}.
\end{aligned} \tag{6.8}$$

We remind that $s_t = x_t^{1/2}$ and that $R_t^{(cl)} = x_t^{1/2} \gamma_t^{(R)}$. Also, the SDE for $I_t^{(cl)} = x_t^{1/2} \gamma_t^{(I)}$ is in the same form as that of $R_t^{(cl)}$ since $\gamma_t^{(I)}$ has the same SDE as $\gamma_t^{(R)}$. Therefore, the system of SDE for the clutter in the coordinate system $\begin{bmatrix} x_t & R_t^{(cl)} & I_t^{(cl)} \end{bmatrix}^\top$ becomes:

$$\begin{bmatrix} \mathrm{d}x_t \\ \mathrm{d}R_t^{(cl)} \\ \mathrm{d}I_t^{(cl)} \end{bmatrix} = \begin{bmatrix} \mathcal{A}(1 - x_t) \\ \left(-\frac{\mathcal{A}+\mathcal{B}}{2} R_t^{(cl)} + \frac{\mathcal{A}R_t^{(cl)}}{2x_t}\left(1 - \frac{1}{2\alpha}\right) \right) \\ \left(-\frac{\mathcal{A}+\mathcal{B}}{2} I_t^{(cl)} + \frac{\mathcal{A}I_t^{(cl)}}{2x_t}\left(1 - \frac{1}{2\alpha}\right) \right) \end{bmatrix} \mathrm{d}t$$
$$+ \begin{bmatrix} \left(\frac{2\mathcal{A}x_t}{\alpha}\right)^{1/2} & 0 & 0 \\ \frac{R_t^{(cl)}}{x_t^{1/2}}\left(\frac{\mathcal{A}}{2\alpha}\right)^{1/2} & \left(\frac{\mathcal{B}x_t}{2}\right)^{1/2} & 0 \\ \frac{I_t^{(cl)}}{x_t^{1/2}}\left(\frac{\mathcal{A}}{2\alpha}\right)^{1/2} & 0 & \left(\frac{\mathcal{B}x_t}{2}\right)^{1/2} \end{bmatrix} \begin{bmatrix} \mathrm{d}W_t^{(x)} \\ \mathrm{d}W_t^{(R)} \\ \mathrm{d}W_t^{(I)} \end{bmatrix}. \qquad (6.9)$$

6.1.2 Homodyned K scattering

In Homodyned K (HK) scattering, the reflectivity of the target is a complex constant added to the reflectivity of the clutter. We denote $\Psi_t^{(HK)}$ the total reflectivity and we have:

$$\Psi_t^{(HK)} = \Psi_c^{(R)} + i\Psi_c^{(I)} + x_t^{1/2}\gamma_t, \qquad (6.10)$$

where $\Psi_c = \Psi_c^{(R)} + i\Psi_c^{(I)}$ is the reflectivity of the target. It is constant in both phase and amplitude. For simplicity, we denote R_t, I_t the real and imaginary parts of $\Psi_t^{(HK)}$. We omit voluntarily the superscript (HK) to lighten the notation.

It is very straightforward to obtain the SDE of R_t and I_t in the case of HK scattering. Indeed, from (6.10) we get:

$$\begin{cases} R_t = \Psi_c^{(R)} + R_t^{(cl)} \, ; \mathrm{d}R_t = \mathrm{d}R_t^{(cl)} \\ I_t = \Psi_c^{(I)} + I_t^{(cl)} \, ; \mathrm{d}I_t = \mathrm{d}I_t^{(cl)}. \end{cases} \qquad (6.11)$$

Therefore, we have directly from equation (6.9):

$$\begin{bmatrix} \mathrm{d}x_t \\ \mathrm{d}R_t \\ \mathrm{d}I_t \end{bmatrix} = \begin{bmatrix} \mathcal{A}(1 - x_t) \\ -\frac{\mathcal{A}+\mathcal{B}}{2}\left(R_t - \Psi_c^{(R)}\right) + \frac{\mathcal{A}\left(R_t - \Psi_c^{(R)}\right)}{2x_t}\left(1 - \frac{1}{2\alpha}\right) \\ -\frac{\mathcal{A}+\mathcal{B}}{2}\left(I_t - \Psi_c^{(I)}\right) + \frac{\mathcal{A}\left(I_t - \Psi_c^{(I)}\right)}{2x_t}\left(1 - \frac{1}{2\alpha}\right) \end{bmatrix} \mathrm{d}t$$
$$+ \begin{bmatrix} \left(\frac{2\mathcal{A}x_t}{\alpha}\right)^{1/2} & 0 & 0 \\ \frac{\left(R_t - \Psi_c^{(R)}\right)}{x_t^{1/2}}\left(\frac{\mathcal{A}}{2\alpha}\right)^{1/2} & \left(\frac{\mathcal{B}x_t}{2}\right)^{1/2} & 0 \\ \frac{\left(I_t - \Psi_c^{(I)}\right)}{x_t^{1/2}}\left(\frac{\mathcal{A}}{2\alpha}\right)^{1/2} & 0 & \left(\frac{\mathcal{B}x_t}{2}\right)^{1/2} \end{bmatrix} \begin{bmatrix} \mathrm{d}W_t^{(x)} \\ \mathrm{d}W_t^{(R)} \\ \mathrm{d}W_t^{(I)} \end{bmatrix}. \qquad (6.12)$$

The SDE obtained for HK scattering, equation (6.12), can also be written in the more compact form:

$$\begin{bmatrix} \mathrm{d}x_t \\ \mathrm{d}R_t \\ \mathrm{d}I_t \end{bmatrix} = \beta_{\Psi_c}^{(HK)}(x_t, R_t, I_t)\,\mathrm{d}t + \Sigma_{\Psi_c}^{(HK)}(x_t, R_t, I_t) \begin{bmatrix} \mathrm{d}W_t^{(x)} \\ \mathrm{d}W_t^{(R)} \\ \mathrm{d}W_t^{(I)} \end{bmatrix}. \qquad (6.13)$$

Vector $\beta_{\Psi_c}^{(HK)}$ and matrix $\Sigma_{\Psi_c}^{(HK)}$ are respectively the drift and volatility in the case of HK scattering.

6.1.3 Generalized K scattering

In Generalized K (GK) scattering, the reflectivity of the target is modulated by the RCS x_t (see [48] p 71):

$$\Psi_t^{(GK)} = \left(\Psi_c^{(R)} + i\Psi_c^{(I)}\right)\eta x_t + x_t^{1/2}\gamma_t, \tag{6.14}$$

where η is a constant coupling factor. Physically, it means that the reflectivity of the target, $\left(\Psi_c^{(R)} + i\Psi_c^{(I)}\right)x_t$, varies proportionally as the number of scatterers which contribute to the clutter. Similarly to section 6.1.2, R_t and I_t denote now the real and imaginary parts of $\Psi_t^{(GK)}$ respectively. From equation (6.14), we get:

$$\begin{cases} R_t = \Psi_c^{(R)}\eta x_t + R_t^{(cl)} \, ; \mathrm{d}R_t = \Psi_c^{(R)}\eta \mathrm{d}x_t + \mathrm{d}R_t^{(cl)} \\ I_t = \Psi_c^{(I)}\eta x_t + I_t^{(cl)} \, ; \mathrm{d}I_t = \Psi_c^{(I)}\eta \mathrm{d}x_t + \mathrm{d}I_t^{(cl)}. \end{cases} \tag{6.15}$$

Again we derive the SDE for R_t and that of I_t will follow immediately. From equations (6.15), (6.9) and (6.3), we get:

$$\begin{aligned} \mathrm{d}R_t &= \left(\eta\Psi_c^{(R)}\mathcal{A}(1-x_t) - \frac{\mathcal{A}+\mathcal{B}}{2}R_t^{(cl)} + \frac{\mathcal{A}R_t^{(cl)}}{2x_t}\left(1-\frac{1}{2\alpha}\right)\right)\mathrm{d}t \\ &+ \left[\frac{R_t^{cl}}{x_t^{1/2}}\left(\frac{\mathcal{A}}{2\alpha}\right)^{1/2} + \eta\Psi_c^{(R)}\left(\frac{2\mathcal{A}x_t}{\alpha}\right)^{1/2} \quad \left(\frac{\mathcal{B}x_t}{2}\right)^{1/2}\right]\begin{bmatrix}\mathrm{d}W_t^{(x)} \\ \mathrm{d}W_t^{(R)}\end{bmatrix}. \end{aligned} \tag{6.16}$$

Replacing $R_t^{(cl)} = R_t - \Psi_c^{(R)}\eta x_t$ in equation (6.16) and applying the same procedure for I_t, we get the following system of coupled SDE that describe GK scattering:

$$\begin{aligned} \begin{bmatrix}\mathrm{d}x_t \\ \mathrm{d}R_t \\ \mathrm{d}I_t\end{bmatrix} &= \begin{bmatrix}\mathcal{A}(1-x_t) \\ \eta\Psi_c^{(R)}\mathcal{A}(1-x_t) + \left(R_t - \Psi_c^{(R)}\eta x_t\right)\left(-\frac{\mathcal{A}+\mathcal{B}}{2} + \frac{\mathcal{A}}{2x_t}\left(1-\frac{1}{2\alpha}\right)\right) \\ \eta\Psi_c^{(I)}\mathcal{A}(1-x_t) + \left(I_t - \Psi_c^{(I)}\eta x_t\right)\left(-\frac{\mathcal{A}+\mathcal{B}}{2} + \frac{\mathcal{A}}{2x_t}\left(1-\frac{1}{2\alpha}\right)\right)\end{bmatrix}\mathrm{d}t \\ &+ \begin{bmatrix}\left(\frac{2\mathcal{A}x_t}{\alpha}\right)^{1/2} & 0 & 0 \\ \frac{R_t-\Psi_c^{(R)}\eta x_t}{x_t^{1/2}}\left(\frac{\mathcal{A}}{2\alpha}\right)^{1/2} + \eta\Psi_c^{(R)}\left(\frac{2\mathcal{A}x_t}{\alpha}\right)^{1/2} & \left(\frac{\mathcal{B}x_t}{2}\right)^{1/2} & 0 \\ \frac{I_t-\Psi_c^{(I)}\eta x_t}{x_t^{1/2}}\left(\frac{\mathcal{A}}{2\alpha}\right)^{1/2} + \eta\Psi_c^{(I)}\left(\frac{2\mathcal{A}x_t}{\alpha}\right)^{1/2} & 0 & \left(\frac{\mathcal{B}x_t}{2}\right)^{1/2}\end{bmatrix}\begin{bmatrix}\mathrm{d}W_t^{(x)} \\ \mathrm{d}W_t^{(R)} \\ \mathrm{d}W_t^{(I)}\end{bmatrix}. \end{aligned} \tag{6.17}$$

The SDE obtained for GK scattering, equation (6.17), can also be written in the more compact form:

$$\begin{bmatrix}\mathrm{d}x_t \\ \mathrm{d}R_t \\ \mathrm{d}I_t\end{bmatrix} = \beta_{\Psi_c}^{(GK)}(x_t, R_t, I_t)\,\mathrm{d}t + \Sigma_{\Psi_c}^{(GK)}(x_t, R_t, I_t)\begin{bmatrix}\mathrm{d}W_t^{(x)} \\ \mathrm{d}W_t^{(R)} \\ \mathrm{d}W_t^{(I)}\end{bmatrix}. \tag{6.18}$$

Vector $\beta_{\Psi_c}^{(GK)}$ and matrix $\Sigma_{\Psi_c}^{(GK)}$ are respectively the drift and volatility in the case of GK scattering.

6.2 Maximum likelihood estimation of Ψ_c

Let $\Psi_c = \Psi_c^{(R)} + i\Psi_c^{(I)}$ be the target constant. We know from section 6.1 that the target reflectivity is simply the target constant Ψ_c for HK scattering, and is the target constant times the RCS and the coupling factor, *i.e.* $\Psi_c \eta x_t$ for GK scattering. In both cases, our aim is to estimate Ψ_c. It is also assumed that there is always a target, since even the absence of target constant to be estimated can be seen as the special case $\Psi_c = 0$.

We assume in this section that we observe three discrete time series:

$$(\tilde{x}, \tilde{R}, \tilde{I}) = \left\{ \left(\tilde{x}_k, \tilde{R}_k, \tilde{I}_k \right), k = 0, 1, \ldots, n \right\},$$

where measurements are made at times t_k. We also assume for simplicity that $\forall k, \ t_k - t_{k-1} = \Delta t$ is a constant.

As explained in chapter 2 section 2.3.1, the maximum likelihood (ML) estimation consists in maximizing the likelihood function with respect to some parameter, namely Ψ_c here. The likelihood can be written:

$$\mathcal{L}\left(\tilde{x}, \tilde{R}, \tilde{I}; \Psi_c \right) = p_{\Psi_c}^{\infty} \prod_{k=1}^{n} p_{\Psi_c}^{(k)}, \tag{6.19}$$

with

$$
\begin{aligned}
p_{\Psi_c}^{\infty} &= p_{\Psi_c}((x_{t_0}, R_{t_0}, I_{t_0}) = (\tilde{x}_0, \tilde{R}_0, \tilde{I}_0)) \\
&= p_{\Psi_c}^{\infty}(\tilde{x}_0, \tilde{R}_0, \tilde{I}_0)
\end{aligned}
\tag{6.20}
$$

and

$$p_{\Psi_c}^{(k)} = p_{\Psi_c}\left((x_{t_k}, R_{t_k}, I_{t_k}) = (\tilde{x}_k, \tilde{R}_k, \tilde{I}_k) \mid (x_{t_{k-1}}, R_{t_{k-1}}, I_{t_{k-1}}) = (\tilde{x}_{k-1}, \tilde{R}_{k-1}, \tilde{I}_{k-1}) \right). \tag{6.21}$$

$p_{\Psi_c}^{\infty}$ is the stationary (asymptotic) distribution of the process (x_t, R_t, I_t). It accounts for the initial value at time t_0 and it is implicit that it is evaluated at $(\tilde{x}_0, \tilde{R}_0, \tilde{I}_0)$. $p_{\Psi_c}^{(k)}$ is the transition probability between times t_{k-1} and t_k. $\mathcal{L}\left(\tilde{x}, \tilde{R}, \tilde{I}; \Psi_c \right)$ is therefore the joint probability (density) of the observed initial value and of the n transitions occuring from t_0 to t_n. We assume that the parameters of the clutter, *i.e.* $\mathcal{A}, \mathcal{B}, C$ and α are known. In that case, for a given observed trajectory, $\mathcal{L}\left(\tilde{x}, \tilde{R}, \tilde{I}; \Psi_c \right)$ depends only on Ψ_c. Maximizing it with respect to Ψ_c yields the estimated target constant $\tilde{\Psi}_c$.

Instead of maximizing directly the likelihood function $\mathcal{L}$, we choose to maximize its logarithm:

$$
\begin{aligned}
l\left(\tilde{x}, \tilde{R}, \tilde{I}; \Psi_c \right) &= \ln\left(\mathcal{L}\left(\tilde{x}, \tilde{R}, \tilde{I}; \Psi_c \right) \right) \\
&= \ln\left(p_{\Psi_c}^{\infty} \right) + \sum_{k=1}^{n} \ln\left(p_{\Psi_c}^{(k)} \right).
\end{aligned}
\tag{6.22}
$$

If $\tilde{x}, \tilde{R}, \tilde{I}$ are fixed, *i.e.* a time series is observed, the necessary conditions of optimality are:

$$
\begin{cases}
\frac{\partial l}{\partial \Psi_c^{(R)}}(\tilde{\Psi}_{c,ML}^{(R)}, \tilde{\Psi}_{c,ML}^{(I)}) = 0 \\
\frac{\partial l}{\partial \Psi_c^{(I)}}(\tilde{\Psi}_{c,ML}^{(R)}, \tilde{\Psi}_{c,ML}^{(I)}) = 0.
\end{cases}
\tag{6.23}
$$

6.2.1 ML estimation of Ψ_c in HK scattering

From equations (6.22) and (6.23), it is obvious that one must know $p_{\Psi_c}^\infty$ and the transition probabilities $p_{\Psi_c}^{(k)}$ to estimate Ψ_c.

Asymptotic distribution

Let $\Psi_t^{(HK)} = \Psi_c^{(R)} + i\Psi_c^{(I)} + x_t^{1/2}\gamma_t$. At fixed RCS $x_t = x$, the random variable (R_t, I_t) is a bivariate Gaussian distribution (see [141] equation (4.12)):

$$p_{\Psi_c}\left((R_t, I_t) = (R, I) \mid x_t = x\right) = \frac{1}{\pi x}\exp\left(-\frac{(R - \Psi_c^{(R)})^2 + (I - \Psi_c^{(I)})^2}{x}\right). \qquad (6.24)$$

In that case, $(R_t^2 + I_t^2)^{1/2}$ follows the so-called Rice distribution. From equation (6.24) and the asymptotic distribution of x_t (*e.g.* equation (4.19)), we get:

$$p_{\Psi_c}^\infty(x, R, I) = \frac{1}{\pi x}\exp\left(-\frac{(R - \Psi_c^{(R)})^2 + (I - \Psi_c^{(I)})^2}{x}\right) \times \frac{\alpha^\alpha x^{\alpha-1} e^{-\alpha x}}{\Gamma(\alpha)}. \qquad (6.25)$$

Approximate transition probabilities

Deriving the transition probabilities requires some tedious calculus. Exact transition probabilities could be derived by solving the Fokker-Planck equation associated with (6.12) (see [120] for a detailed account of the Fokker-Planck equation). In practice, it is rarely possible to solve it analytically. Instead, if the time difference Δt between t_{k-1} and t_k is small enough, one can approximate the transition probability by a multivariate Gaussian distribution. The starting point is Euler-Maruyama's scheme applied to the SDE (6.13):

$$\begin{bmatrix}\Delta x_t \\ \Delta R_t \\ \Delta I_t\end{bmatrix} = \beta_{\Psi_c}^{(HK)}(x_t, R_t, I_t)\,\Delta t + \Sigma_{\Psi_c}^{(HK)}(x_t, R_t, I_t)\begin{bmatrix}\Delta W_t^{(x)} \\ \Delta W_t^{(R)} \\ \Delta W_t^{(I)}\end{bmatrix}$$

$$\Leftrightarrow\quad \begin{bmatrix}x_{t+\Delta t} \\ R_{t+\Delta t} \\ I_{t+\Delta t}\end{bmatrix} = \begin{bmatrix}x_t \\ R_t \\ I_t\end{bmatrix} + \beta_{\Psi_c}^{(HK)}(x_t, R_t, I_t)\,\Delta t + \Sigma_{\Psi_c}^{(HK)}(x_t, R_t, I_t)\,\Delta t^{1/2}\begin{bmatrix}n_x \\ n_R \\ n_I\end{bmatrix}, \quad (6.26)$$

where $\begin{bmatrix}n_x & n_R & n_I\end{bmatrix}^T$ is a vector of independent standard Gaussian random variables.
It follows that:

$$p_{\Psi_c}^{(k)} \approx \frac{1}{(2\pi)^{3/2}\left|\sigma_{\Psi_c}^{(HK)}\Delta t\right|^{1/2}}\exp\left(-\frac{1}{2}(v_k - \mu_k)^T\left(\sigma_{\Psi_c}^{(HK)}\Delta t\right)^{-1}(v_k - \mu_k)\right), \qquad (6.27)$$

with

$$\begin{cases}v_k = \begin{bmatrix}\tilde{x}_k & \tilde{R}_k & \tilde{I}_k\end{bmatrix}^\top \\ \mu_k = \begin{bmatrix}\tilde{x}_{k-1} & \tilde{R}_{k-1} & \tilde{I}_{k-1}\end{bmatrix}^\top + \beta_{\Psi_c}^{(HK)}\left(\tilde{x}_{k-1}, \tilde{R}_{k-1}, \tilde{I}_{k-1}\right)\Delta t \\ \sigma_{\Psi_c}^{(HK)} = \Sigma_{\Psi_c}^{(HK)}\Sigma_{\Psi_c}^{(HK)\,\top}.\end{cases} \qquad (6.28)$$

Indeed, to get the transition probability, we fix $\begin{bmatrix} x_t & R_t & I_t \end{bmatrix}^\top = \begin{bmatrix} \tilde{x}_{k-1} & \tilde{R}_{k-1} & \tilde{I}_{k-1} \end{bmatrix}^\top$. Thus, the only random part in the right hand side of equation (6.26) is the Gaussian vector $\begin{bmatrix} n_x & n_R & n_I \end{bmatrix}^\top$.

$\sigma_{\Psi_c}^{(HK)}$ is referred to as the squared volatility. It is a positive definite symmetric matrix in the non degenerate case. We have directly:

$$\sigma_{\Psi_c}^{(HK)} = \begin{bmatrix} \frac{2\mathcal{A}x_t}{\alpha} & \frac{\mathcal{A}(R_t-\Psi_c^{(R)})}{\alpha} & \frac{\mathcal{A}(I_t-\Psi_c^{(I)})}{\alpha} \\ \frac{\mathcal{A}(R_t-\Psi_c^{(R)})}{\alpha} & \frac{\mathcal{A}}{2\alpha}\frac{(R_t-\Psi_c^{(R)})^2}{x_t} + \frac{\mathcal{B}x_t}{2} & \frac{\mathcal{A}}{2\alpha}\frac{(R_t-\Psi_c^{(R)})(I_t-\Psi_c^{(I)})}{x_t} \\ \frac{\mathcal{A}(I_t-\Psi_c^{(I)})}{\alpha} & \frac{\mathcal{A}}{2\alpha}\frac{(R_t-\Psi_c^{(R)})(I_t-\Psi_c^{(I)})}{x_t} & \frac{\mathcal{A}}{2\alpha}\frac{(I_t-\Psi_c^{(I)})^2}{x_t} + \frac{\mathcal{B}x_t}{2} \end{bmatrix}. \tag{6.29}$$

We have very simply that:

$$\left| \sigma_{\Psi_c}^{(HK)}\Delta t \right| = \left| \Sigma_{\Psi_c}^{(HK)} \right|^2 \Delta t^3 = \frac{\Delta t^3 \mathcal{A}\mathcal{B}^2 x_t^3}{2\alpha}. \tag{6.30}$$

We can then invert $\sigma_{\Psi_c}^{(HK)}$ and we get after some calculations (see appendix D.1):

$$\left(\sigma_{\Psi_c}^{(HK)}\Delta t \right)^{-1} = \begin{bmatrix} \frac{(R_t-\Psi_c^{(R)})^2+(I_t-\Psi_c^{(I)})^2}{2\mathcal{B}\Delta t x_t^3} + \frac{\alpha}{2\mathcal{A}\Delta t x_t} & -\frac{R_t-\Psi_c^{(R)}}{\mathcal{B}\Delta t x_t^2} & -\frac{I_t-\Psi_c^{(I)}}{\mathcal{B}\Delta t x_t^2} \\ -\frac{R_t-\Psi_c^{(R)}}{\mathcal{B}\Delta t x_t^2} & \frac{2}{\mathcal{B}\Delta t x_t} & 0 \\ -\frac{I_t-\Psi_c^{(I)}}{\mathcal{B}\Delta t x_t^2} & 0 & \frac{2}{\mathcal{B}\Delta t x_t} \end{bmatrix}. \tag{6.31}$$

Note that in equation (6.27), $\left(\sigma_{\Psi_c}^{(HK)}\Delta t \right)^{-1}$ must be evaluated at $t = t_{k-1}$, i.e. $[x_t, R_t, I_t] = [\tilde{x}_{k-1}, \tilde{R}_{k-1}, \tilde{I}_{k-1}]$ respectively in equation (6.29). Using (6.22), (6.25) and (6.27), we have:

$$l\left(\tilde{x}, \tilde{R}, \tilde{I}; \Psi_c \right) = \ln\left(\frac{\alpha^\alpha \tilde{x}_0^{\alpha-1} e^{-\alpha \tilde{x}_0}}{\pi \tilde{x}_0 \Gamma(\alpha)} \right) - \frac{(\tilde{R}_0 - \Psi_c^{(R)})^2 + (\tilde{I}_0 - \Psi_c^{(I)})^2}{\tilde{x}_0}$$

$$-n\ln\left((2\pi)^{3/2} \left| \sigma_{\Psi_c}^{(HK)}\Delta t \right|^{1/2} \right) + \sum_{k=1}^{n} \Phi_k \tag{6.32}$$

with

$$\Phi_k = -\frac{1}{2}(v_k - \mu_k)^T \left(\sigma_{\Psi_c}^{(HK)}\Delta t \right)^{-1} (v_k - \mu_k). \tag{6.33}$$

If we express the first optimality condition in (6.23), we get after some calculations (see appendix D.1):

$$\frac{\partial l}{\partial \Psi_c^{(R)}}(\tilde{\Psi}_{c,ML}^{(R)}, \tilde{\Psi}_{c,ML}^{(I)}) = 0$$

$$\Leftrightarrow \quad -\frac{2\left(\tilde{\Psi}_{c,ML}^{(R)} - \tilde{R}_0 \right)}{\tilde{x}_0} + \sum_{k=1}^{n} \frac{\tilde{\Psi}_{c,ML}^{(R)} - \tilde{R}_{k-1}}{\mathcal{B}\Delta t \tilde{x}_{k-1}} \left(-\frac{w_k^2}{2\tilde{x}_{k-1}^2} - 2\gamma_k^2 + \frac{2\gamma_k w_k}{\tilde{x}_{k-1}} \right)$$

$$+ \sum_{k=1}^{n} \frac{\tilde{R}_k - \tilde{R}_{k-1}}{\mathcal{B}\Delta t \tilde{x}_{k-1}} \left(2\gamma_k - \frac{w_k}{\tilde{x}_{k-1}} \right) = 0, \tag{6.34}$$

with

$$\begin{cases} w_k = \tilde{x}_k - \tilde{x}_{k-1} - \mathcal{A}\Delta t(1 - \tilde{x}_{k-1}) \\ \gamma_k = \frac{\mathcal{A}+\mathcal{B}}{2} - \frac{\mathcal{A}}{2\tilde{x}_{k-1}}\left(1 - \frac{1}{2\alpha} \right). \end{cases} \tag{6.35}$$

We see that equation (6.34) depends only on x_t and the real part R_t of the reflectivity. It gives a condition on $\tilde{\Psi}_{c,ML}^{(R)}$ only. It is then straightforward to get the estimation of $\Psi_c^{(R)}$:

$$\tilde{\Psi}_{c,ML}^{(R)} = \frac{-\frac{2\tilde{R}_0}{\tilde{x}_0} + \sum_{k=1}^n \frac{\tilde{R}_{k-1}}{\mathcal{B}\Delta t \tilde{x}_{k-1}}\left(-\frac{w_k^2}{2\tilde{x}_{k-1}^2} - 2\gamma_k^2 + \frac{2\gamma_k w_k}{\tilde{x}_{k-1}}\right) - \sum_{k=1}^n \frac{\tilde{R}_k - \tilde{R}_{k-1}}{\mathcal{B}\Delta t \tilde{x}_{k-1}}\left(2\gamma_k - \frac{w_k}{\tilde{x}_{k-1}}\right)}{\sum_{k=1}^n \frac{1}{\mathcal{B}\Delta t \tilde{x}_{k-1}}\left(-\frac{w_k^2}{2\tilde{x}_{k-1}^2} - 2\gamma_k^2 + \frac{2\gamma_k w_k}{\tilde{x}_{k-1}}\right) - \frac{2}{\tilde{x}_0}}$$

(6.36)

The estimator for $\Psi_c^{(I)}$ follows the same equation with $\tilde{I}$ replacing $\tilde{R}$. Note that w_k and γ_k are common to both the estimators of $\Psi_c^{(R)}$ and $\Psi_c^{(I)}$.

6.2.2   ML estimation of Ψ_c in GK scattering

As seen below, the estimator for Ψ_c in the GK scattering case is obtained by the same procedure as for HK scattering.

Asymptotic distribution

Let $\Psi_t^{(GK)} = \left(\Psi_c^{(R)} + i\Psi_c^{(I)}\right)\eta x_t + x_t^{1/2}\gamma_t$. At fixed RCS $x_t = x$, the random variable (R_t, I_t) is a bivariate Gaussian distribution (see [141] equation (4.12)):

$$p_{\Psi_c}\left((R_t, I_t) = (R, I) \mid x_t = x\right) = \frac{1}{\pi x}\exp\left(-\frac{(R - \Psi_c^{(R)}\eta x)^2 + (I - \Psi_c^{(I)}\eta x)^2}{x}\right). \quad (6.37)$$

From equation (6.37) and the asymptotic distribution of x_t (equation (4.19)), we get:

$$p_{\Psi_c}^\infty(x, R, I) = \frac{1}{\pi x}\exp\left(-\frac{(R - \Psi_c^{(R)}\eta x)^2 + (I - \Psi_c^{(I)}\eta x)^2}{x}\right) \times \frac{\alpha^\alpha x^{\alpha-1} e^{-\alpha x}}{\Gamma(\alpha)}. \quad (6.38)$$

Approximate transition probabilities

To derive approximate transition probabilities, we dodge the complications of the Fokker-Planck equation again and compute Gaussian approximations. If we apply Euler-Maruyama's scheme to (6.18), we get:

$$\begin{bmatrix} \Delta x_t \\ \Delta R_t \\ \Delta I_t \end{bmatrix} = \beta_{\Psi_c}^{(GK)}(x_t, R_t, I_t)\,\Delta t + \Sigma_{\Psi_c}^{(GK)}(x_t, R_t, I_t)\begin{bmatrix} \Delta W_t^{(x)} \\ \Delta W_t^{(R)} \\ \Delta W_t^{(I)} \end{bmatrix}$$

$$\Leftrightarrow \quad \begin{bmatrix} x_{t+\Delta t} \\ R_{t+\Delta t} \\ I_{t+\Delta t} \end{bmatrix} = \begin{bmatrix} x_t \\ R_t \\ I_t \end{bmatrix} + \beta_{\Psi_c}^{(GK)}(x_t, R_t, I_t)\,\Delta t + \Sigma_{\Psi_c}^{(GK)}(x_t, R_t, I_t)\,\Delta t^{1/2}\begin{bmatrix} n_x \\ n_R \\ n_I \end{bmatrix}, \quad (6.39)$$

where $\begin{bmatrix} n_x & n_R & n_I \end{bmatrix}^\top$ is a vector of independent standard Gaussian random variables.

It follows that:

$$p_{\Psi_c}^{(k)} \approx \frac{1}{(2\pi)^{3/2}\left|\sigma_{\Psi_c}^{(GK)}\Delta t\right|^{1/2}}\exp\left(-\frac{1}{2}(v_k - \mu_k)^\top\left(\sigma_{\Psi_c}^{(GK)}\Delta t\right)^{-1}(v_k - \mu_k)\right), \quad (6.40)$$

with

$$\begin{cases} v_k = \begin{bmatrix} \tilde{x}_k & \tilde{R}_k & \tilde{I}_k \end{bmatrix}^\top \\ \mu_k = \begin{bmatrix} \tilde{x}_{k-1} & \tilde{R}_{k-1} & \tilde{I}_{k-1} \end{bmatrix}^\top + \beta_{\Psi_c}^{(GK)}\left(\tilde{x}_{k-1}, \tilde{R}_{k-1}, \tilde{I}_{k-1}\right)\Delta t \\ \sigma_{\Psi_c}^{(GK)} = \Sigma_{\Psi_c}^{(GK)}\Sigma_{\Psi_c}^{(GK)\top}. \end{cases} \quad (6.41)$$

Again, to get the transition probability, we fixed $\begin{bmatrix} x_t & R_t & I_t \end{bmatrix}^\top = \begin{bmatrix} \tilde{x}_{k-1} & \tilde{R}_{k-1} & \tilde{I}_{k-1} \end{bmatrix}^\top$. Thus, the only random part in the right hand side of equation (6.39) is the Gaussian vector $\begin{bmatrix} n_x & n_R & n_I \end{bmatrix}^\top$.

The squared volatility $\sigma_{\Psi_c}^{(GK)}$ is also a positive definite symmetric matrix in the non degenerate case. We have directly:

$$\sigma_{\Psi_c}^{(GK)} = \begin{bmatrix} \dfrac{2\mathcal{A}x_t}{\alpha} & \dfrac{\mathcal{A}(R_t + \Psi_c^{(R)}\eta x_t)}{\alpha} & \dfrac{\mathcal{A}(I_t + \Psi_c^{(I)}\eta x_t)}{\alpha} \\ \dfrac{\mathcal{A}(R_t + \Psi_c^{(R)}\eta x_t)}{\alpha} & \dfrac{\mathcal{A}}{2\alpha}\dfrac{(R_t + \Psi_c^{(R)}\eta x_t)^2}{x_t} + \dfrac{\mathcal{B}x_t}{2} & \dfrac{\mathcal{A}}{2\alpha}\dfrac{(R_t + \Psi_c^{(R)}\eta x_t)(I_t + \Psi_c^{(I)}\eta x_t)}{x_t} \\ \dfrac{\mathcal{A}(I_t + \Psi_c^{(I)}\eta x_t)}{\alpha} & \dfrac{\mathcal{A}}{2\alpha}\dfrac{(R_t + \Psi_c^{(R)}\eta x_t)(I_t + \Psi_c^{(I)}\eta x_t)}{x_t} & \dfrac{\mathcal{A}}{2\alpha}\dfrac{(I_t + \Psi_c^{(I)}\eta x_t)^2}{x_t} + \dfrac{\mathcal{B}x_t}{2} \end{bmatrix}. \quad (6.42)$$

We have again that:

$$\left| \sigma_{\Psi_c}^{(GK)}\Delta t \right| = \left| \Sigma_{\Psi_c}^{(GK)} \right|^2 \Delta t^3 = \frac{\Delta t^3 \mathcal{A}\mathcal{B}^2 x_t^3}{2\alpha}. \quad (6.43)$$

If we invert $\sigma_{\Psi_c}^{(GK)}$ we get after some calculations (see appendix D.2):

$$\left(\sigma_{\Psi_c}^{(GK)}\Delta_t \right)^{-1} = \begin{bmatrix} \dfrac{(R_t + \Psi_c^{(R)}\eta x_t)^2 + (I_t + \Psi_c^{(I)}\eta x_t)^2}{2\mathcal{B}\Delta t x_t^3} + \dfrac{\alpha}{2\mathcal{A}\Delta t x_t} & -\dfrac{R_t + \Psi_c^{(R)}\eta x_t}{\mathcal{B}\Delta t x_t^2} & -\dfrac{I_t + \Psi_c^{(I)}\eta x_t}{\mathcal{B}\Delta t x_t^2} \\ -\dfrac{R_t + \Psi_c^{(R)}\eta x_t}{\mathcal{B}\Delta t x_t^2} & \dfrac{2}{\mathcal{B}\Delta t x_t} & 0 \\ -\dfrac{I_t + \Psi_c^{(I)}\eta x_t}{\mathcal{B}\Delta t x_t^2} & 0 & \dfrac{2}{\mathcal{B}\Delta t x_t} \end{bmatrix}. \quad (6.44)$$

In equation (6.40), $\left(\sigma_{\Psi_c}^{(GK)}\Delta t \right)^{-1}$ must be evaluated at $t = t_{k-1}$, i.e. $\begin{bmatrix} x_t & R_t & I_t \end{bmatrix} = \begin{bmatrix} \tilde{x}_{k-1} & \tilde{R}_{k-1} & \tilde{I}_{k-1} \end{bmatrix}$ in equation (6.42). Using equations (6.22), (6.38) and (6.40), we have:

$$l\left(\tilde{x}, \tilde{R}, I; \tilde{\Psi}_c \right) = \ln\left(\frac{\alpha^\alpha \tilde{x}_0^{\alpha-1} e^{-\alpha \tilde{x}_0}}{\pi \tilde{x}_0 \Gamma(\alpha)} \right) - \frac{(\tilde{R}_0 - \Psi_c^{(R)}\eta \tilde{x}_0)^2 + (\tilde{I}_0 - \Psi_c^{(I)}\eta \tilde{x}_0)^2}{\tilde{x}_0}$$
$$-n\ln\left((2\pi)^{3/2}\left| \sigma_{\Psi_c}^{(HK)}\Delta t \right|^{1/2} \right) + \sum_{k=1}^{n}\Phi_k \quad (6.45)$$

with

$$\Phi_k = -\frac{1}{2}(v_k - \mu_k)^\top \left(\sigma_{\Psi_c}^{(HK)}\Delta t \right)^{-1}(v_k - \mu_k). \quad (6.46)$$

If we express the first optimality condition in (6.23), we get after some calculations (see appendix D.2):

$$\frac{\partial l}{\partial \Psi_c^{(R)}}(\tilde{\Psi}_{c,ML}^{(R)}, \tilde{\Psi}_{c,ML}^{(I)}) = 0$$

$$\Leftrightarrow \quad -\frac{2\eta\left(\tilde{\Psi}_{c,ML}^{(R)}\eta \tilde{x}_0 - \tilde{R}_0 \right)}{\tilde{x}_0} + \sum_{k=1}^{n}\lambda_k^{(1)} + \tilde{\Psi}_c^{(R)}\sum_{k=1}^{n}\lambda_k^{(2)} = 0, \quad (6.47)$$

with

$$\lambda_k^{(1)} = \frac{-w_k^2 \eta \tilde{R}_{k-1}}{2\mathcal{B}\Delta t \tilde{x}_{k-1}^2} - \frac{2}{\mathcal{B}\Delta t \tilde{x}_{k-1}} \left(\tilde{R}_k - \tilde{R}_{k-1} - \gamma_k \Delta t \tilde{R}_{k-1} \right)$$
$$\times \left(-\eta\mathcal{A}\Delta t(1 - \tilde{x}_{k-1}) + \gamma_k \Delta t \eta \tilde{x}_{k-1} \right) + \frac{w_k}{\mathcal{B}\Delta t \tilde{x}_{k-1}^2} \eta \tilde{x}_{k-1} \left(\tilde{R}_k - \tilde{R}_{k-1} - \gamma_k \Delta t \tilde{R}_{k-1} \right)$$
$$+ \frac{w_k}{\mathcal{B}\Delta t \tilde{x}_{k-1}^2} \tilde{R}_{k-1} \left(-\eta\mathcal{A}\Delta t(1 - \tilde{x}_{k-1}) + \gamma_k \Delta t \eta \tilde{x}_{k-1} \right) \tag{6.48}$$

and

$$\lambda_k^{(2)} = \frac{-w_k^2 \eta^2}{2\mathcal{B}\Delta t \tilde{x}_{k-1}} - \frac{2}{\mathcal{B}\Delta t \tilde{x}_{k-1}} \left(-\eta\mathcal{A}\Delta t(1 - \tilde{x}_{k-1}) + \gamma_k \Delta t \eta \tilde{x}_{k-1} \right)$$
$$\times \left(-\eta\mathcal{A}\Delta t(1 - \tilde{x}_{k-1}) + \gamma_k \Delta t \eta \tilde{x}_{k-1} \right)$$
$$+ \frac{2w_k \eta}{\mathcal{B}\Delta t \tilde{x}_{k-1}} \left(-\eta\mathcal{A}\Delta t(1 - \tilde{x}_{k-1}) + \gamma_k \Delta t \eta \tilde{x}_{k-1} \right). \tag{6.49}$$

Equation (6.47) depends only on x_t and the real part R_t of the reflectivity. It gives a condition on $\tilde{\Psi}_{c,ML}^{(R)}$ only. It is straightforward to get the estimation of $\Psi_c^{(R)}$:

$$\tilde{\Psi}_{c,ML}^{(R)} = \frac{-\frac{2\eta \tilde{R}_0}{\tilde{x}_0} - \sum_{k=1}^n \lambda_k^{(1)}}{-2\eta^2 + \sum_{k=1}^n \lambda_k^{(2)}}. \tag{6.50}$$

The estimator for $\Psi_c^{(I)}$ follows the same equation with $\tilde{I}$ replacing $\tilde{R}$. Note that w_k and γ_k are common to both the estimators of $\Psi_c^{(R)}$ and $\Psi_c^{(I)}$.

6.2.3 Notational remarks

As in chapter 5 for the ML estimators of $\mathcal{A}$ and $\mathcal{B}$, we used here an observed *trajectory* (one realization) $\left(\tilde{x}, \tilde{R}, \tilde{I} \right) = \left\{ \left(\tilde{x}_k, \tilde{R}_k, \tilde{I}_k \right), k = 0, 1, \ldots, n \right\}$ to estimate $\Psi_c^{(R)}$ and $\Psi_c^{(I)}$. It was again for presentational clarity but it means that rigorously, $\tilde{\Psi}_{c,ML}^{(R)}$ and $\tilde{\Psi}_{c,ML}^{(I)}$ are estimates, not estimators. However, considering that the trajectory is random, we can consider that $\tilde{\Psi}_{c,ML}^{(R)}$ and $\tilde{\Psi}_{c,ML}^{(I)}$ are truly estimators.

6.3 Performance of the ML estimation

6.3.1 A simple estimator for Ψ_c

As seen in the previous section, estimating Ψ_c with ML is straightforward but involves quite heavy expressions, at least as compared to the ergodicity-based estimator. This very simple estimator arises naturally. Indeed from equations (6.10) and (6.14) we get:

$$\begin{cases} \mathbb{E}\left[\Psi_c^{(HK)} \right] = \mathbb{E}\left[\Psi_c + x_t^{1/2} \gamma_t \right] = \Psi_c + \mathbb{E}[x_t^{1/2}]\mathbb{E}[\gamma_t] = \Psi_c \\ \mathbb{E}\left[\Psi_c^{(GK)} \right] = \mathbb{E}\left[\Psi_c \eta x_t + x_t^{1/2} \gamma_t \right] = \eta\Psi_c \mathbb{E}[x_t] + \mathbb{E}[x_t^{1/2}]\mathbb{E}[\gamma_t] = \eta\Psi_c. \end{cases} \tag{6.51}$$

We have used that x_t and γ_t are independent, and that $\mathbb{E}[x_t] = 1$ and $\mathbb{E}[\gamma_t] = 0$ (see chapter 3 section 3.3). We assume that the reflectivity is ergodic, or more precisely we assume that:

$$\mathbb{E}\left[\Psi_t^{(HK)} \right] = \lim_{T \to +\infty} \frac{1}{T} \int_0^T \tilde{\Psi}_t^{(HK)} \mathrm{d}t \tag{6.52}$$

for any particular trajectory $\tilde{\Psi}_t$ of the reflectivity, in which case:

$$\Psi_c = \lim_{T \to +\infty} \frac{1}{T} \int_0^T \tilde{\Psi}_t^{(HK)} \mathrm{d}t \approx \frac{1}{t_n} \int_0^{t_n} \tilde{\Psi}_t^{(HK)} \mathrm{d}t \approx \frac{1}{n\Delta t} \sum_{i=0}^{n-1} \tilde{\Psi}_i^{(HK)} \Delta t. \qquad (6.53)$$

Of course, the same holds in the GK scattering case, such that the ergodicity-based estimators (or moment) yield:

$$\begin{cases} \tilde{\Psi}_{c,e} = \frac{1}{n\Delta t} \sum_{i=0}^{n-1} \tilde{\Psi}_i^{(HK)} \Delta t & (HK\,scattering) \\ \tilde{\Psi}_{c,e} = \frac{1}{\eta n\Delta t} \sum_{i=0}^{n-1} \tilde{\Psi}_i^{(GK)} \Delta t & (GK\,scattering). \end{cases} \qquad (6.54)$$

6.3.2   Numerical experiments

In this section, we compare the ML estimator to the ergodicity-based estimator which serves as a reference. What we really want is to test numerically whether it is relevant to use the ML estimator, whose formalism is more heavy. To assess the abilities of the ML and ergodicity-based estimators, we simulate many trajectories of $\begin{bmatrix} x_t & R_t & I_t \end{bmatrix}^\top$ with a given Ψ_c and then try to retrieve it. To do so, we solve numerically the SDE (6.12) and (6.17) using Euler-Maruyama's scheme for R_t, I_t and Milstein's scheme for x_t (see [62] and chapter 2 section 2.2.3). The Euler-Maruyama scheme is not used for x_t because it raises numerical issues. Indeed, it could generate negative values, which is absurd since the RCS is always positive. The trajectories are simulated with parameters relevant for the sea surface, which is a special case of random medium. As in chapters 4 and 5, we take $\mathcal{A} = 1$ Hz, $\mathcal{B} = 100$ Hz, $\alpha = 1$ and $\Delta t = 0.001$ s. We set $\eta = 1$ for the GK scattering simulations.

The detection of a coherent scatterer is naturally more challenging if the coherent scatterer has a power similar to the clutter. We explore the dependence of the estimator performance to both the trajectory duration and target intensity (power). We define the target intensity simply as its squared modulus:

$$|\Psi_c|^2 = \Psi_c^{(R)\,2} + \Psi_c^{(I)\,2} \qquad (6.55)$$

By rotational symmetry and for simplicity, we also set $\Psi_c^{(I)} = 0$ such that all the target power goes into the real part. We denote $\tilde{\Psi}_{c,i}^{(R)}$ the estimated $\Psi_c^{(R)}$ from the i-th trajectory, from either $\tilde{\Psi}_{c,ML}$ (maximum likelihood) or $\tilde{\Psi}_{c,e}$ ("ergodicity"), and M the number of trajectories. We assess the performance of the estimators by computing their estimation bias $b(\tilde{\Psi}_c^{(R)})$ and variance $\sigma^2(\tilde{\Psi}_c^{(R)})$:

$$b(\tilde{\Psi}_c^{(R)}) = \frac{1}{M} \sum_{i=1}^{M} \left(\tilde{\Psi}_{c,i}^{(R)} - \bar{\Psi}_c^{(R)} \right), \qquad (6.56)$$

$$\sigma^2(\tilde{\Psi}_c^{(R)}) = \frac{1}{M-1} \sum_{i=1}^{M} \left(\tilde{\Psi}_{c,i}^{(R)} - \bar{\Psi}_c^{(R)} \right)^2, \qquad (6.57)$$

where $\bar{\Psi}_c^{(R)}$ is the average of the $\tilde{\Psi}_{c,i}^{(R)}$.

Performance in HK scattering

In the case of HK scattering (equation (6.10)), we numerically solve the SDE (6.12) for known clutter parameters $\mathcal{A}, \mathcal{B}, \alpha$ and target $\Psi_c = \Psi_c^{(R)}$. We then try to retrieve $\Psi_c^{(R)}$ using equations

([6.36]) and ([6.54]) which correspond respectively to the ML and ergodicity-based estimators. We explore the dependency of the estimator performances to both the trajectory duration (from 0.1 s to 10 s) and target intensity (from 0.1 to 100). When the duration is explored, the target intensity is set to 10, and when the target intensity is explored, the duration is set to 1 s. The relative squared estimation bias and the estimation variance are plotted in figure 6.1. They are simply the squared bias and variance defined in equation ([6.56]) normalized by the target intensity. We observe that both decrease quickly as the trajectory duration or target intensity increase. This is easily understandable since a longer trajectory carries more information, and a stronger target dominates more strongly the overall reflectivity. However, the squared bias is orders of magnitude smaller than the variance and is therefore negligible. It is also what explains its apparent noisiness. If we now compare the ergodicity-based and ML estimators based on the variance, we notice a slight advantage for the ML estimator for low target intensity.

Figure 6.1: Relative estimation squared bias and variance of the target in HK scattering for both the ergodicity and ML estimators. 1000 trajectories are computed, with $\mathcal{A} = 1$ Hz, $\mathcal{B} = 100$ Hz and $\alpha = 1$. Up: dependence to trajectory duration with $\Psi_c = \sqrt{10}$. Down: dependence to target intensity with a duration of 1 s.

We represent in figure 6.2 a scatter plot of all the estimated $\Psi_c^{(R)}$ for 1000 trajectories. There is no striking difference between the ergodicity and ML clouds, unlike the case of GK scattering.

Performance in GK scattering

In the case of GK scattering (equation ([6.14])), we numerically solve the SDE ([6.17]) for known clutter parameters $\mathcal{A}, \mathcal{B}, \alpha$, coupling η and target $\Psi_c = \Psi_c^{(R)}$. We then try to retrieve $\Psi_c^{(R)}$
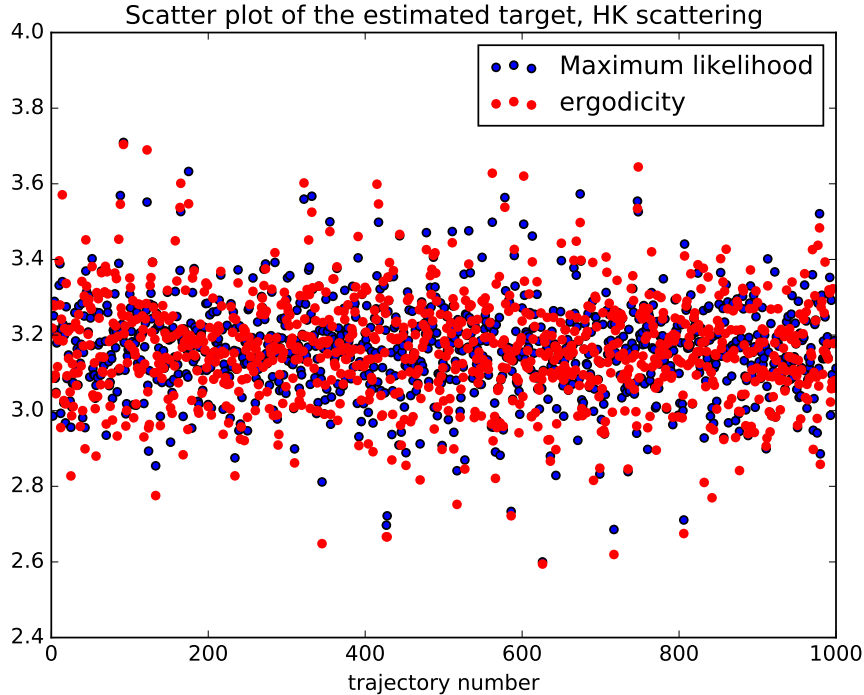
Figure 6.2: Scatter plot of the estimated target for both the ergodicity and ML estimators in the case of HK scattering. 1000 trajectories of duration 1 s are computed, with $\mathcal{A} = 1$ Hz, $\mathcal{B} = 100$ Hz, $\alpha = 1$ and $\Psi_c^{(R)} = \sqrt{10}$.

using equations (6.50) and (6.54) which correspond respectively to the ML and ergodicity-based estimators. We explore the dependency of the estimators performances to both the trajectory duration (from 0.1 s to 10 s) and target intensity (from 0.1 to 100). Again, when the duration is explored, the target intensity is set to 10, and when the target intensity is explored, the duration is set to 1 s. The relative squared estimation bias and the estimation variance are plotted in figure 6.3. We also observe that the variance decreases quickly as the trajectory duration or target intensity increase. The bias is very noiselike but it is again orders of magnitude smaller than the variance so we shall ignore it. Unlike HK scattering, there is this time a large difference between the performance of the ergodicity and ML estimator. The ML estimator outperforms the ergodicity estimator by an order of magnitude on average (less for small durations and target intensity and vice versa).

We represent in figure 6.4 a scatter plot of all the estimated $\Psi_c^{(R)}$ for 1000 trajectories. There is a striking difference between the ergodicity and ML clouds. There are both centered correctly, but the ML cloud is much narrower than the ergodicity one, in accordance with the estimation variances. There is a second difference that we should mention: the ML estimator has the drawback of generating some outliers (on the order of 0.5%). These outliers have been filtered out for the variance calculations of figure 6.3. We noticed that they arise when during the trajectory, x_t goes very close to 0. Of course, it is not physically possible to have a vanishing RCS. However, we will keep the outliers in the discussion (section 6.4), since they merely increase the false alarm rate by a small percentage.
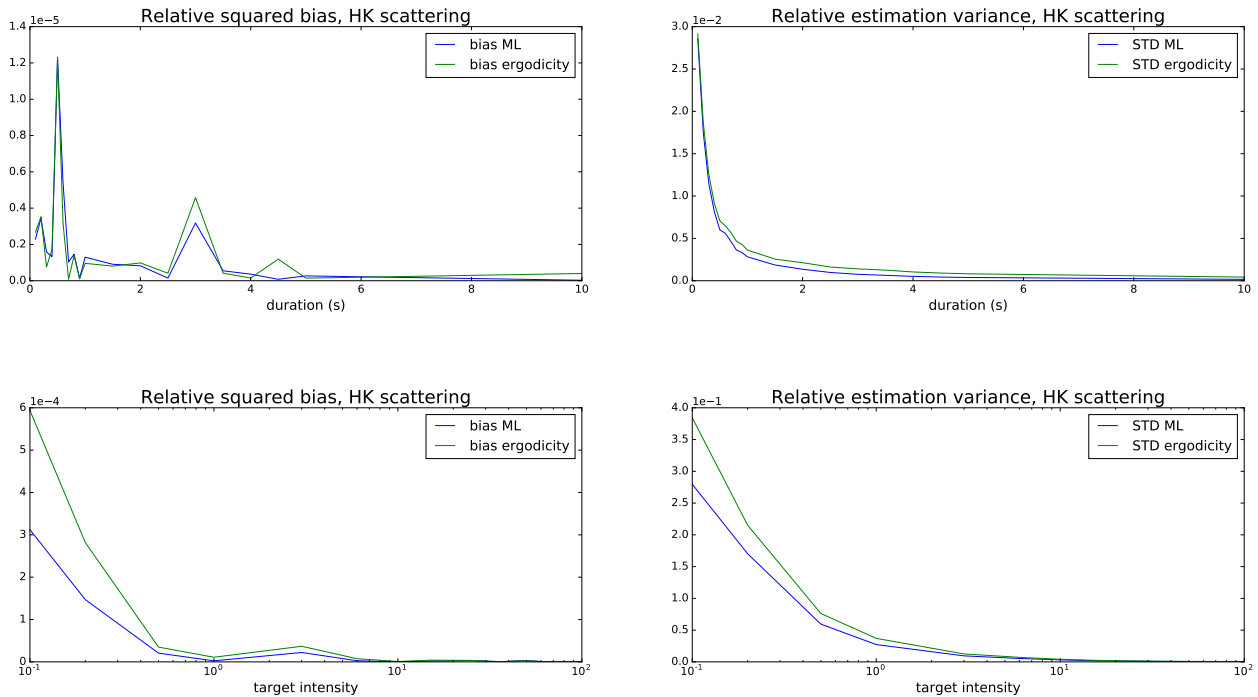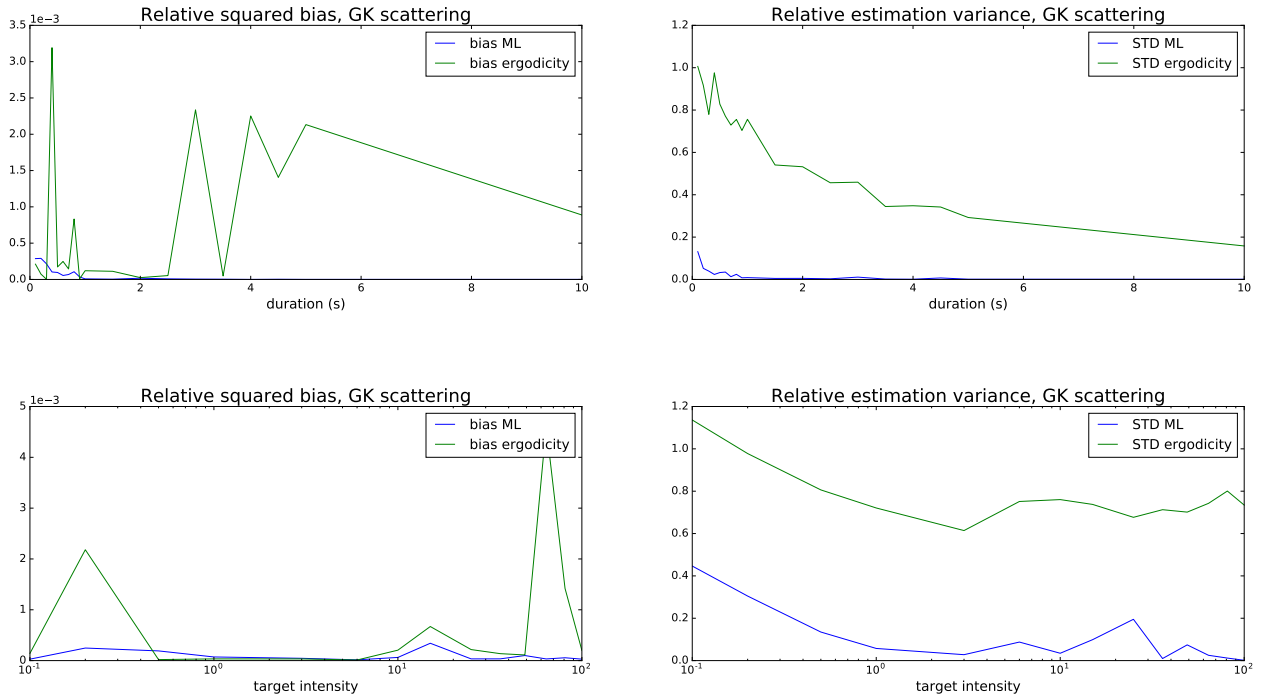
154

Figure 6.3: Relative estimation squared bias and variance of the target in GK scattering for both the ergodicity and ML estimators. 1000 trajectories are computed, with $\mathcal{A} = 1$ Hz, $\mathcal{B} = 100$ Hz, $\alpha = 1$ and $\eta = 1$. Up: dependence to trajectory duration with $\Psi_c = \sqrt{10}$. Down: dependence to target intensity with a duration of 1 s..

Comparison between HK and GK scattering

In HK scattering, the target reflectivity Ψ_c is simply added to the random medium reflectivity (clutter) $\Psi_t = x_t^{1/2} \gamma_t$. There is no interaction between the target and random medium. Retrospectively, it is then natural that the transition probabilities do not 'really' depend on the target and that ML estimation does not present any advantage. More precisely, we mean that the probability of increasing or decreasing by some amount during time interval Δt is independent on the target strengh in HK scattering. In that case, ML estimation is not better than the ergodicity-based estimator, except in the case of a low intensity target.

In GK scattering, the target reflectivity is modulated by the RCS to give $\Psi_c \eta x_t$ which is added to Ψ_t. There is an interaction between the target and random medium, and in this case the transition probabilities 'really' depend on the target. As a result, ML estimation is relevant since every single transition depends on Ψ_c. Comparing figure 6.1 and figure 6.3, we also see that the estimation variance is greater for GK than HK scattering, which is also a result of the target being modulated by x_t in GK scattering.

Performance as a function of the sampling frequency in GK scattering

We now want to assess the influence of the sampling frequency. So far, we fixed the time step to $\Delta t = 10^{-3}$ s. We just saw that the ML estimator significantly outperforms the ergodicity-based estimator in the case of GK scattering. Since this is due to the information about Ψ_c contained in every single transition, it is natural to think that the more transitions, the better
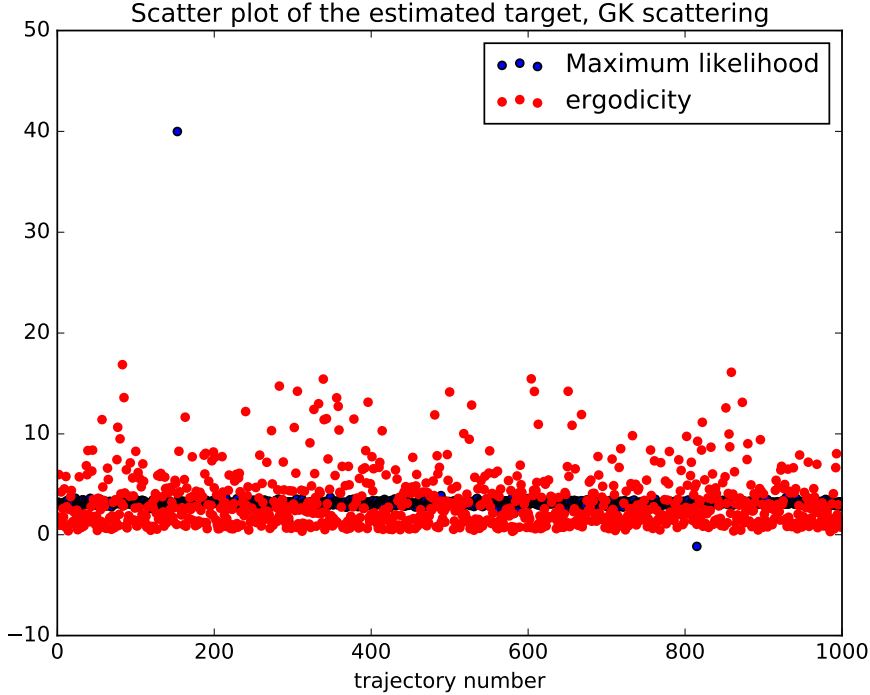
Figure 6.4: Scatter plot of the estimated target for both the ergodicity and ML estimators. 1000 trajectories of duration 1 s are computed, with $\mathcal{A} = 1$ Hz, $\mathcal{B} = 100$ Hz, $\alpha = 1$, $\eta = 1$ and $\Psi_c^{(R)} = \sqrt{10}$.

the performance. To test this idea, we carried numerical simulations and estimated Ψ_c for a range of sampling timesteps from 10^{-5} s to 10^{-1} s. For each timestep and value of $\mathcal{A}$, 10000 trajectories are computed with $\Psi_c = \sqrt{10}$ (only for GK scattering), and Ψ_c is estimated by ML. The estimation variance is then computed. The results are represented in figure 6.5. We observe as expected that the variance decreases as the sampling time step decreases (increase of the sampling frequency) most likely to an asymptotic lower bound. However, it decreases very slightly compared to the changes relative to the duration (compare figure 6.5 and figure 6.3). In GK scattering, there is information about the target in every transition due to the fact that x_t modulates the target constant Ψ_c in GK scattering. Therefore, as well as changing the timestep, we also change how fast x_t evolves, by tuning $\mathcal{A}$. Based on our simulation results, it seems that increasing the dynamics of x_t, *i.e.* increasing $\mathcal{A}$, leads to a decrease in the estimation variance.

6.4 Discussion

We propose a first approach to the detection problem (instead of estimation) considering the distribution of the estimated parameter $\tilde{\Psi}_c$. For simplicity, let us assume that the target constant is real-valued *i.e.* $\Psi_c = \Psi_c^{(R)}$. We denote $p_{\Psi_c}\left(\tilde{\Psi}_c^{(R)} = x\right)$ and $p_0\left(\tilde{\Psi}_c^{(R)} = x\right)$ the distributions of $\tilde{\Psi}_c^{(R)}$ with and without a target Ψ_c respectively. $\tilde{\Psi}_c^{(R)}$ denotes any estimator of $\Psi_c^{(R)}$, for example $\tilde{\Psi}_{c,ML}^{(R)}$ or $\tilde{\Psi}_{c,e}^{(R)}$. For target detection, one can evaluate the likelihood
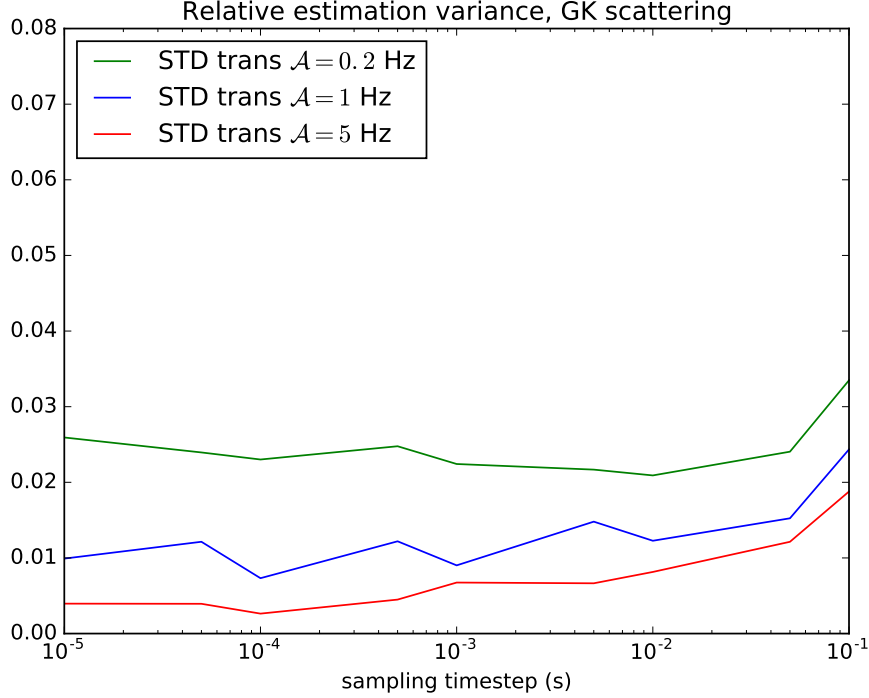
Figure 6.5: Relative estimation variance of the target in GK scattering for the ML estimators. 10000 trajectories are computed, with $\mathcal{B} = 100$ Hz, $\alpha = 1$ and $\eta = 1$. The variance is plotted as a function of the sampling time step.

ratio:

$$\Lambda(\tilde{\Psi}_c^{(R)}) = \frac{p_{\Psi_c}\left(\tilde{\Psi}_c^{(R)} = x\right)}{p_0\left(\tilde{\Psi}_c^{(R)} = x\right)}. \tag{6.58}$$

Our method for estimating Ψ_c does not provide the distributions at the numerator and denominator for computing Λ. Instead, one can compute numerical distributions by estimating Ψ_c on many trajectories. The approach is identical to that of section 6.4. Examples of numerical distributions obtained with 10000 trajectories are represented in figure 6.6 (left) for two target constants: $\Psi_c = 1$ and $\Psi_c = \sqrt{0.1}$. In addition, we represent the Gaussian distributions with the numerical means and variances. To first order, the curves fit reasonably well. We observe that the separation between the no target and target distributions is much smaller for $\Psi_c = \sqrt{0.1}$.

If we denote $\sigma_{\Psi_c}, \sigma_0$ and m_{Ψ_c}, m_0 the numerical standard deviations and means with and without a target, we have:

$$\Lambda(\tilde{\Psi}_c^{(R)}) \approx \frac{\frac{1}{\sqrt{2\pi}\sigma_{\Psi_c}}e^{-(\tilde{\Psi}_c^{(R)} - m_{\Psi_c})/2\sigma_{\Psi_c}^2}}{\frac{1}{\sqrt{2\pi}\sigma_0}e^{-(\tilde{\Psi}_c^{(R)} - m_0)/2\sigma_0^2}}. \tag{6.59}$$

By deciding that there is a target if Λ exceeds the threshold λ_T and varying the threshold, we compute a range of probabilities of detection and probabilities of false alarms P_D and P_{FA} and get a ROC curve. More precisely, 10000 trajectories of duration 1 s are numerically computed with fixed $\Psi_c = \Psi_c^{(R)} > 0$ (target). The i-th trajectory gives an estimated target $\tilde{\Psi}_{c,i}^{(R)}$. m_{Ψ_c} and σ_{Ψ_c} are then computed from the set $\{\tilde{\Psi}_{c,i}^{(R)}, i = 1, 2, \ldots, 10000\}$ of estimated
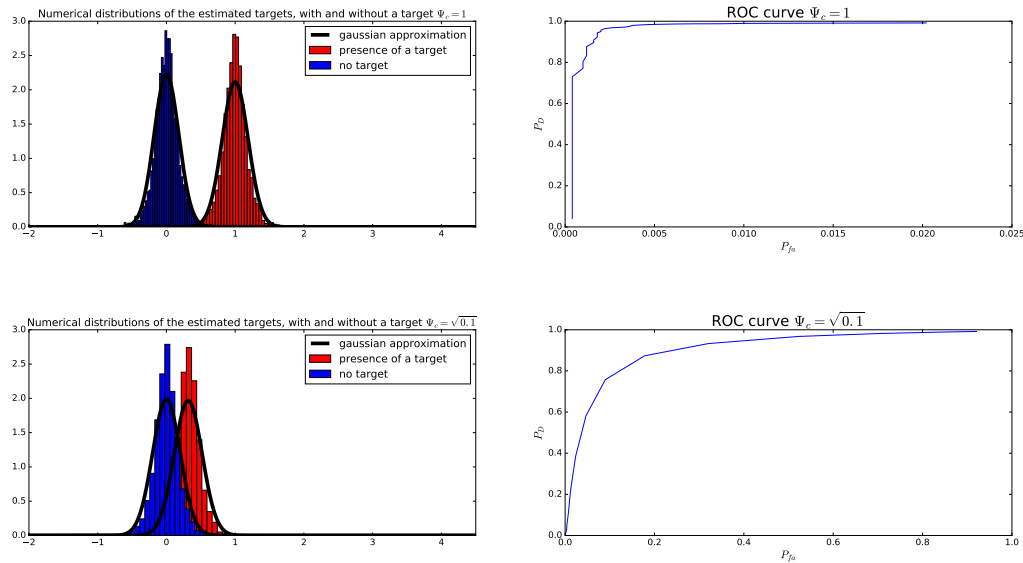
Figure 6.6: Numerical distributions of $\tilde{\Psi}_c$ with and without a target, and numerical ROC curves based on the distribution of $\tilde{\Psi}_c$. 10000 trajectories of duration 1 s are computed, with $\mathcal{A} = 1$ Hz, $\mathcal{B} = 100$ Hz, $\alpha = 1$, $\eta = 1$. Up: $\Psi_c = 1$; Down: $\Psi_c = \sqrt{0.1}$.

$\Psi_c^{(R)}$. A similar procedure is carried to compute numerically m_0 and σ_0 (no target). Finally, to compute the probability of detection at fixed threshold, we count how many trajectories generated with a target $\Psi_c = \Psi_c^{(R)} > 0$ are such that their associated $\tilde{\Psi}_{c,i}^{(R)}$ verifies the condition $\Lambda(\tilde{\Psi}_{c,i}^{(R)}) > \lambda_T$. A similar procedure is done for the probability of false alarm.

The ROC curves with $\Psi_c = 1$ and $\Psi_c = \sqrt{0.1}$ are represented in figure 6.6 (right). Notice that the range covered by the x-axis depends on the target. It is manifest that the ROC curve is much better with $\Psi_c = 1$ ('strong' target) than with $\Psi_c = \sqrt{0.1}$ ('weak' target).

The reader may object that the Gaussian distributions do not fit so well the numerical distributions. We wish to emphasize that we chose it just for illustrative purposes. More generally, the objective of this discussion is to show that the estimation of the target Ψ_c indeed leads to target detection, which is discussed again in chapter 7.

6.5 Conclusion

In this chapter, we derived the SDE for the in-phase (R_t) and quadrature phase (I_t) components of the reflectivity of a random medium (for example the sea surface). Two models of target have been considered: HK scattering and GK scattering. Approximate transition probabilities have been derived for small time steps using Euler-Maruyama scheme. We used these approximate transition probabilities for maximum likelihood estimation of the coherent scatterer (target) constant. In both HK and GK scattering, the estimator is given by an explicit analytical formula.

Numerical simulations have been carried out using constants relevant for radar scattering of the sea surface. The estimation bias is negligible and the estimation variance is small compared to the target constant. We observed the intuitive result that the estimation variance decreases as the duration, or target intensity, or sampling frequency increases. We

showed that mostly for GK scattering is the maximum likelihood estimator significantly better than the very simple ergodicity-based estimator. Only in this case, do the transition probabilities really (*i.e.* not a simple translation) depend on the target constant due to target/clutter interaction. For HK scattering, the maximum likelihood estimator is better than the ergodicity-based estimator in the case of a weak target. Lastly, we discussed how our estimation of the target parameters leads to target detection, ROC curves etc.

We think that it would be valuable to study more precisely the performance of the estimation in terms of the quantity of information about the target contained in the time series (depending on duration and sampling frequency). This quantity of information would also depend on the parameters $\mathcal{A}$, $\mathcal{B}$ and α, which control the dynamics of the clutter. Finally, HK and GK scattering are two examples of target reflectivity models. It is possible that based on physical considerations, one would choose a more adapted model, maybe a combination of HK and GK scattering or a model where the target constant is not modulated by x_t but by a different process. Our approach could then be extended to such models as long as the dynamics are expressed by stochastic differential equations.

Chapter 7

Non-observability of x_t, bayesian estimation and target detection

In chapter 5, we presented the estimation of the clutter parameters of Field's model: $\mathcal{A}$, $\mathcal{B}$ and α. In chapter 6, we presented the estimation of the target parameters $\Psi_c^{(R)}$, $\Psi_c^{(I)}$. Under the assumption that α is known and that the RCS x_t is observed (and also that there is no target), we estimated $\mathcal{A}$ and $\mathcal{B}$ by maximum likelihood (ML). Under the assumption that $\mathcal{A}$, $\mathcal{B}$ and α are known, and that x_t is observed, we also estimated $\Psi_c^{(R)}$ and $\Psi_c^{(I)}$ by ML. In this chapter, we address the issue of the non-observability of x_t. Indeed, real data recorded by a coherent radar yield time series of Ψ_t, but its factorization into the product of $\Psi_t = x_t^{1/2}\gamma_t$ is not readily available. Estimating x_t from observable data becomes necessary to be able to apply the results of the previous chapters. The first purpose of this chapter is to solve the problem of estimating the clutter parameters without observing x_t (sections 7.1 and 7.2).

The second purpose of this chapter is to introduce new tools and directions for future work on the use of Field's model to describe the sea clutter. As such, the second part of this chapter (sections 7.3 and 7.4) is exploratory. We introduce two new topics: bayesian statistics for sea clutter, and target detection. First, notwithstanding the frequentist point of view adopted so far for estimating the clutter and target parameters, is it possible at low cost to adopt the bayesian point of view for estimation? Second, in chapter 6 the target was always present and constant. Shall we now consider that the target appears for some time and then disappears (time dependence), and can we define algorithms which can "scan" a time series of the complex reflectivity, and react if there is an anomaly compared to the normal situation (sea clutter only)?

This chapter is organized as follows. In section 7.1, we derive estimators for x_t. Corresponding estimators for γ_t are immediate. In section 7.2, we merge all the results obtained for estimating the parameters $\mathcal{A}$, $\mathcal{B}$ and α, as well as the results for estimating x_t to propose a chain of estimators applicable to the data that would really be observed in practice. In section 7.3, we introduce bayesian estimation of the clutter parameters $\mathcal{A}$, $\mathcal{B}$ and α. In section 7.4, we define two algorithms for target detection: one based on spot volatility, and one on the likelihood of the observed time series. Section 7.5 concludes.

161

7.1 Estimation of x_t and γ_t

In practice, with a coherent radar, one can observe the complex process Ψ_t. We remind that the complex reflectivity of the random medium (the clutter) is:

$$\Psi_t = x_t^{1/2}\gamma_t = x_t^{1/2}\gamma_t^{(R)} + ix_t^{1/2}\gamma_t^{(I)} = R_t + iI_t. \tag{7.1}$$

Only $R_t = x_t^{1/2}\gamma_t^{(R)}$ and $I_t = x_t^{1/2}\gamma_t^{(I)}$ are observed, but they both enclose the two processes x_t and γ_t and it is not trivial to disentangle them. Nevertheless, in chapter 5, we assumed that we had the decomposition of Ψ_t in terms of x_t and γ_t. Then, x_t was used to estimate $\mathcal{A}$ and γ_t was used to estimate $\mathcal{B}$. Also, in chapter 6, we assumed that x_t was observed to estimate the target parameters.

We remind from chapter 3 that x_t is the radar cross section (RCS) of the sea surface. If x_t is set to one (its mean value since Ψ_t is normalized), then $\Psi_t = \gamma_t$, *i.e.* the complex reflectivity reduces to the speckle. It is possible to modulate the average power of the speckle by multiplying it by a slower positive process, for example $x_t^{1/2}$ in Field's model. Mathematically speaking, there is no reason to think that x_t should be slower than γ_t. However, based on physical considerations, we asserted it from chapter 4 on, by taking $\mathcal{A} = 1$ Hz (dynamics of x_t) and $\mathcal{B} = 100$ Hz (dynamics of the speckle) in most of our simulations. This difference of timescales between x_t and γ_t is the key to estimate x_t.

The problem of estimating x_t in the case of clutter only (no target, corresponding to equation (7.1)) from a discrete time series $\{\tilde{\Psi}_k, k = 1, 2, \ldots, n\}$ has been adressed in [44]. It is not very clear what the authors do (since for example there is no equation where the estimator of x_t is clearly defined), but they seem to estimate x_t based on the spot volatility of the phase θ_t, where $\Psi_t = z_t^{1/2}e^{i\theta_t}$. For a stochastic process X_t solving:

$$dX_t = \mu(X_t)dt + \Sigma(X_t)dW_t, \tag{7.2}$$

the spot volatility is simply $\sigma(X_t) = \Sigma^2(X_t)$, *i.e.* the squared volatility. For example the spot volatility of x_t, which solves the first SDE of equation (6.3), is $2\frac{\mathcal{A}}{\alpha}x_t$. In [44], the authors use averaging over a sliding window, and the optimal size of the window (in the sense that it minimizes the mean square error between the true and estimated x_t) is proposed as a function of $\mathcal{A}$, α and Δt. [45] is a work similar to [44] but it considers an inverse gamma distribution for x_t.

In this section, we propose several estimators for x_t, but only one is retained, namely $\breve{x}_t$ (see below). An estimator for x_t naturally leads to an estimator for γ_t, here: $\breve{\gamma}_t = \frac{\Psi_t}{\breve{x}_t^{1/2}}$. Since the estimators for x_t imply that we smooth some signal with a sliding window, we study numerically what the optimal size of the window is as a function of $\mathcal{A}$ and α.

7.1.1 Heuristic estimators for x_t

$\bar{x}_t$ estimator

To propose an estimator for x_t, we first introduce notations and basic results. Let

$$\Delta_k\Psi = \Psi_{t_k} - \Psi_{t_{k-1}}. \tag{7.3}$$

We have:

$$\begin{aligned}
\Delta_k\Psi &= x_{t_k}^{1/2}\gamma_{t_k} - x_{t_{k-1}}^{1/2}\gamma_{t_{k-1}} \\
&= \left(x_{t_k}^{1/2}\gamma_{t_k}^{(R)} - x_{t_{k-1}}^{1/2}\gamma_{t_{k-1}}^{(R)}\right) + i\left(x_{t_k}^{1/2}\gamma_{t_k}^{(I)} - x_{t_{k-1}}^{1/2}\gamma_{t_{k-1}}^{(I)}\right).
\end{aligned} \tag{7.4}$$

We get:

$$|\Delta_k \Psi|^2 = \left(x_{t_k}^{1/2} \gamma_{t_k}^{(R)} - x_{t_{k-1}}^{1/2} \gamma_{t_{k-1}}^{(R)} \right)^2 + \left(x_{t_k}^{1/2} \gamma_{t_k}^{(I)} - x_{t_{k-1}}^{1/2} \gamma_{t_{k-1}}^{(I)} \right)^2.$$

If $\Delta t = t_k - t_{k-1}$ is small compared to the time scale of x_t (which holds for example for $\mathcal{A} = 1$ Hz and $\Delta t = 10^{-3}$ s), then $x_{t_{k-1}} \approx x_{t_k}$ and we have:

$$|\Delta_k \Psi|^2 \approx x_{t_k} \left[\left(\gamma_{t_k}^{(R)} - \gamma_{t_{k-1}}^{(R)} \right)^2 + \left(\gamma_{t_k}^{(I)} - \gamma_{t_{k-1}}^{(I)} \right)^2 \right]. \tag{7.5}$$

By Euler-Maruyama's scheme applied to the SDE:

$$\begin{cases} \mathrm{d}\gamma_t^{(R)} = -\frac{1}{2}\mathcal{B}\gamma_t^{(R)}\mathrm{d}t + \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\mathrm{d}W_t^{(R)} \\ \mathrm{d}\gamma_t^{(I)} = -\frac{1}{2}\mathcal{B}\gamma_t^{(I)}\mathrm{d}t + \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\mathrm{d}W_t^{(I)}, \end{cases} \tag{7.6}$$

it holds:

$$\begin{cases} \Delta_k\gamma^{(R)} \approx -\frac{1}{2}\mathcal{B}\gamma_t^{(R)}\Delta t + \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\Delta_k W^{(R)} \\ \Delta_k\gamma^{(I)} \approx -\frac{1}{2}\mathcal{B}\gamma_t^{(R)}\Delta t + \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\Delta_k W^{(I)}. \end{cases} \tag{7.7}$$

Under the hypothesis that Δt is small, the volatility term dominates the drift term and we get:

$$\begin{cases} \Delta_k\gamma^{(R)} \approx \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\Delta_k W^{(R)} = \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}n_k^{(R)}\Delta t^{1/2} \\ \Delta_k\gamma^{(I)} \approx \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\Delta_k W^{(I)} = \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}n_k^{(I)}\Delta t^{1/2}, \end{cases} \tag{7.8}$$

where the $n_k^{(R)}$ and $n_k^{(I)}$ are independent normal random variables such that for all k, $n_k^{(R)} \sim \mathcal{N}(0,1)$ and $n_k^{(I)} \sim \mathcal{N}(0,1)$. From equations (7.5) and (7.8), we get:

$$|\Delta_k \Psi|^2 = x_{t_k} \left[\frac{\mathcal{B}}{2}n_k^{(R)\,2}\Delta t + \frac{\mathcal{B}}{2}n_k^{(I)\,2}\Delta t \right]. \tag{7.9}$$

After these preliminary calculi, we now set t and consider the estimation of x_t. We take time window Δ_t containing t (or preferably centered at t) and make the assumption that for all $k \in \Delta_t$ (more rigourosly for all k such that $t_k \in \Delta_t$), we have $x_{t_k} \approx x_t$. If we average the squared increments of Ψ_t over the window, we then get:

$$\frac{1}{N}\sum_{k\in\Delta_t} |\Delta_k\Psi|^2 \approx x_t \frac{\mathcal{B}\Delta t}{2} \left(\frac{1}{N}\sum_{k\in\Delta_t} n_k^{(R)\,2} + \frac{1}{N}\sum_{k\in\Delta_t} n_k^{(I)\,2} \right) \approx \mathcal{B}\Delta t x_t, \tag{7.10}$$

by the law of large numbers. N is the number of elements in the window. Thus, we obtain the following estimator for x_t:

$$\bar{x}_t = \frac{1}{\mathcal{B}\Delta t N} \sum_{k\in\Delta_t} |\Delta_k\Psi|^2. \tag{7.11}$$

If we develop a bit equation (7.11), we get:

$$\begin{aligned} \bar{x}_t &= \frac{1}{\mathcal{B}\Delta t N} \sum_{k\in\Delta_t} \left[(R_{t_k} - R_{t_{k-1}})^2 + (I_{t_k} - I_{t_{k-1}})^2 \right] \tag{7.12} \\ &= \frac{1}{\mathcal{B}} \left[\frac{1}{T}\sum_{k\in\Delta_t} (R_{t_k} - R_{t_{k-1}})^2 + \frac{1}{T}\sum_{k\in\Delta_t} (I_{t_k} - I_{t_{k-1}})^2 \right] \\ &= \frac{1}{\mathcal{B}} \left(\sigma_t^{RR} + \sigma_t^{II} \right) + \epsilon. \end{aligned}$$

σ_t^{RR} and σ_t^{II} are the RR and II coefficients of the *squared*-volatility of the process $\begin{bmatrix} x_t & R_t & I_t \end{bmatrix}^\top$ (see chapter 6 section 6.1.1). Here, R_t and I_t are for the real and imaginary parts of the sea clutter only. ϵ is a residual in replacing $\frac{1}{T}\sum_{k\in\Delta_t}(R_{t_k}-R_{t_{k-1}})^2$ and $\frac{1}{T}\sum_{k\in\Delta_t}(I_{t_k}-I_{t_{k-1}})^2$ by σ_t^{RR} and σ_t^{II}. It could be large but it will tend to zero as $\Delta_t \to 0$ and $\Delta t \to 0$ (spot volatility estimation, see [8]). From equations (6.9) and (7.12), we get:

$$
\begin{aligned}
\bar{x}_t &= \frac{1}{\mathcal{B}}\left[\frac{\mathcal{A}}{2\alpha x_t}R_t^2 + \frac{\mathcal{B}x_t}{2} + \frac{\mathcal{A}}{2\alpha x_t}I_t^2 + \frac{\mathcal{B}x_t}{2}\right] + \epsilon \\
\Leftrightarrow \bar{x}_t &= x_t + \frac{\mathcal{A}}{2\mathcal{B}\alpha x_t}\left[R_t^2 + I_t^2\right] + \epsilon.
\end{aligned}
\tag{7.13}
$$

As explained, ϵ goes to zero as Δ_t and Δt go to zero, but $\frac{\mathcal{A}}{2\mathcal{B}\alpha x_t}\left[R_t^2 + I_t^2\right]$ has no reason to vanish. Our estimator is therefore not consistent. However, this non-vanishing term is expected to be small since $\frac{\mathcal{A}}{\mathcal{B}} \ll 1$.

$\hat{x}_t$ estimator

A more simple way to estimate x_t is to notice that (see chapter 3 section 3.3):

$$
\mathbb{E}\left[u|\gamma_t^{(R)}|^2\right] = u.
\tag{7.14}
$$

Let us take again a window Δ_t centered at t. If $|\Delta_t|$ (size of the window) is small compared to the decorrelation time of x_t ($\approx 1/\mathcal{A}$) and large compared to the decorrelation time of γ_t ($\approx 1/\mathcal{B}$), then $x_s \approx x_t$ for all $s \in \Delta_t$. We then have:

$$
\frac{1}{N}\sum_{k\in\Delta_t}|\gamma_{t_k}|^2 \approx 1,
\tag{7.15}
$$

by application of the law of large numbers and knowing that $\mathbb{E}[|\gamma_{t_k}|^2] = 1$. Therefore, we have:

$$
\frac{1}{N}\sum_{k\in\Delta_t}z_{t_k} = \frac{1}{N}\sum_{k\in\Delta_t}x_{t_k}|\gamma_{t_k}|^2 \approx x_t\frac{1}{N}\sum_{k\in\Delta_t}|\gamma_{t_k}|^2 \approx x_t.
\tag{7.16}
$$

Thus, we have the following estimator for x_t:

$$
\hat{x}_t = \frac{1}{N}\sum_{k\in\Delta_t}z_{t_k},
\tag{7.17}
$$

i.e. x_t is estimated by a local average of the intensity z_t.

7.1.2 Estimation of $\mathcal{B}$ based on the increments of Ψ_t

To estimate x_t with $\bar{x}_t$, we need to know $\mathcal{B}$. However, to estimate $\mathcal{B}$, γ_t is needed, and thus x_t is needed in the first place. We could use the estimator $\hat{x}_t$, estimate $\mathcal{B}$ from $\hat{\gamma}_t = \Psi_t/\hat{x}_t^{1/2}$, and then use this estimate to compute $\bar{x}_t$. We are not enthusiastic about this solution because two windows are required, one for $\hat{x}_t$ and one for $\bar{x}_t$ (they have no reason to be the same). Instead, there is a much more elegant solution. We remind that equation (7.11) reads:

$$
\bar{x}_t = \frac{1}{\mathcal{B}\Delta t N}\sum_{k\in\Delta_t}|\Delta_k\Psi|^2.
\tag{7.18}
$$

It is quite remarkable that the solution of the problem comes from the problem itself! Indeed, since $\bar{x}_t$ is an estimator of x_t, we can expect that:

$$\mathbb{E}[\bar{x}_t] = \mathbb{E}[x_t] = 1. \tag{7.19}$$

Of course, the first equality is not quite true at the view of equation (7.13) (even asymptotically as the sampling frequency goes to infinity). We admit it though because our numerical simulations showed that the error is small compared to one. We define $\bar{X}_t$ by:

$$\bar{X}_t = \frac{1}{\Delta t N} \sum_{k \in \Delta} |\Delta_k \Psi|^2 = \mathcal{B}\bar{x}_t. \tag{7.20}$$

It approximately holds that:

$$\mathbb{E}[\bar{X}_t] = \mathcal{B}\mathbb{E}[\bar{x}_t] = \mathcal{B}. \tag{7.21}$$

If we assume that $\bar{X}_t$ is ergodic and that $\bar{X}_t$ is computed at times t_i for $i = 1, 2, \ldots, m$ (by centering the averaging window successively over the t_i), we have:

$$\mathbb{E}[\bar{X}_t] = \lim_{T \to +\infty} \frac{1}{T} \int_0^T \bar{X}_r \mathrm{d}r \approx \frac{1}{m\Delta t} \sum_{i=1}^m \bar{X}_{t_i} \Delta t = \frac{1}{m} \sum_{i=1}^m \bar{X}_{t_i}. \tag{7.22}$$

We obtain the following estimator for $\mathcal{B}$:

$$
\begin{aligned}
\tilde{\mathcal{B}}_\Psi &= \frac{1}{m} \sum_{i=1}^m \bar{X}_{t_i} \\
\Leftrightarrow \tilde{\mathcal{B}}_\Psi &= \frac{1}{m\Delta t N} \sum_{i=1}^m \sum_{k \in \Delta_i} |\Delta_k \Psi|^2,
\end{aligned}
\tag{7.23}
$$

where Δ_i is a window centered at time t_i. Because ergodicity is used, the estimator for $\mathcal{B}$ in equation (7.23) is fundamentally different from the ML estimator in chapter 5 (equation (5.33)) and the volatility-based estimator in appendix A.1. Indeed, it requires long trajectories to work since ergodicity is used, and it is based on observable data, while the ML and volatility-based estimators require γ_t. The performance of $\tilde{\mathcal{B}}_\Psi$ is therefore not comparable to that of $\tilde{\mathcal{B}}_{ML}$ (maximum likelihood estimator from chapter 5). However, we do numerical simulations to estimate its estimation bias and standard deviation with 300 s and 1200 s long trajectories. For each value of $\mathcal{B}$ in $\{10, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000\}$ Hz, 1000 trajectories of Ψ_t are simulated with $\mathcal{A} = 1$, $\alpha = 1$ and $\Delta t = 0.001$ s (simulation and estimation time step). $\mathcal{B}$ is estimated with equation (7.23) and the estimation bias and standard deviation are computed. What N shall we choose? We anticipate on the results of section 7.1.4 and particularly on equation (7.27). For $\mathcal{A} = 1$ Hz, $\alpha = 1$ and $\Delta t = 0.001$ s, we get $N_{opt} \approx 100$. The results of the simulations are represented in figure 7.1. The bias is not negligible: it increases as $\mathcal{B}$ increases. However, it does not depend on the duration of the trajectories. For its part, the standard deviation is reduced by a factor two when we go from 300 to 1200 s.
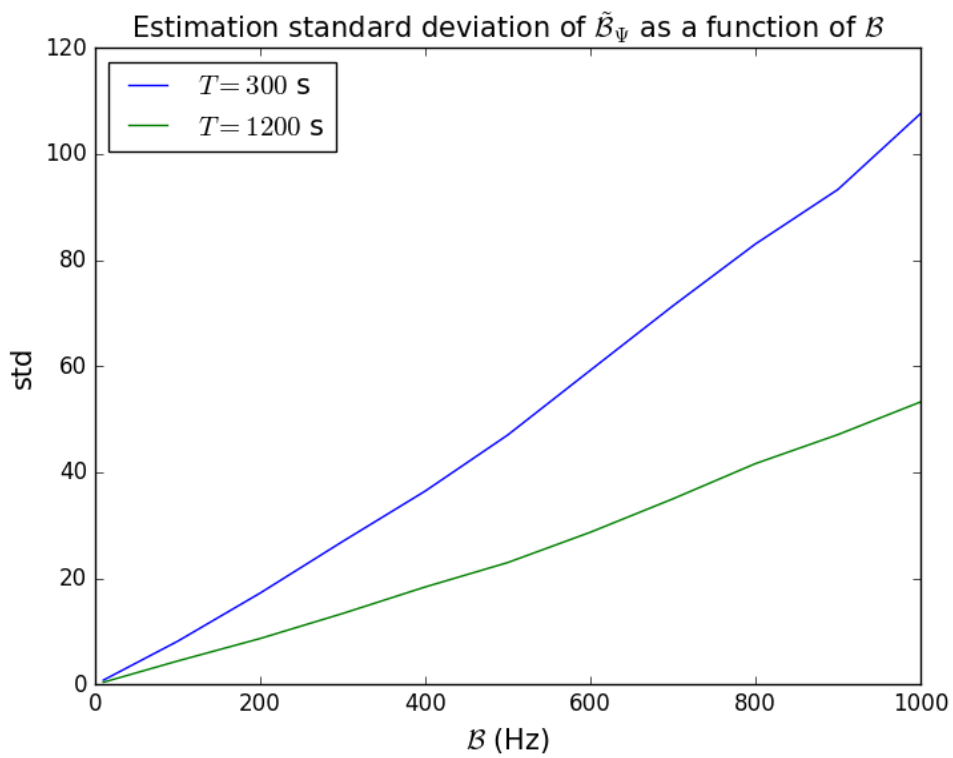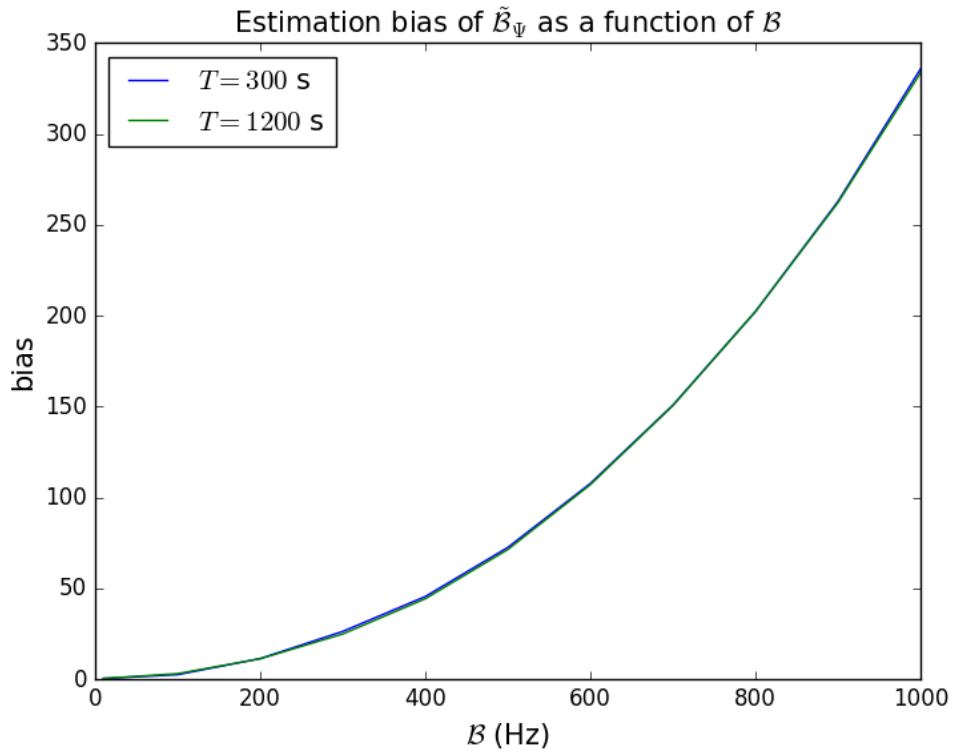
Figure 7.1: Estimation bias (up) and standard deviation (down) of $\tilde{\mathcal{B}}_\Psi$ as a function of true $\mathcal{B}$, for trajectories of duration $T = 300$ s and $T = 1200$ s.

7.1.3   Numerical assessment of the estimators for x_t

We do numerical simulations to compare the estimators of x_t: $\bar{x}_t$ given by equation (7.11) and $\hat{x}_t$ given by equation (7.17). We also define a new estimator, which is almost identical to $\bar{x}_t$, only with the difference that $\tilde{\mathcal{B}}_\Psi$ is used instead of $\mathcal{B}$:

$$\breve{x}_t = \frac{1}{\tilde{\mathcal{B}}_\Psi \Delta t N} \sum_{k \in \Delta_t} |\Delta_k \Psi|^2. \tag{7.24}$$

Of course, by construction:

$$\frac{1}{m} \sum_{i=1}^{m} \breve{x}_{t_i} = 1. \tag{7.25}$$

The purpose is to evaluate $\bar{x}_t$, $\hat{x}_t$ and $\breve{x}_t$, keeping in mind that in practice only $\hat{x}_t$ and $\breve{x}_t$ are computable from real data. From the simulations, we can compute $\bar{x}_t$ since we know the real $\mathcal{B}$ used to simulate the trajectories.

   We simulate 10 trajectories of $\Psi_t = x_t^{1/2} \gamma_t$ using Milstein's scheme for x_t and Euler-Maruyama's scheme for γ_t. We set $\mathcal{A} = 1$ Hz, $\alpha = 1$ and $\mathcal{B} = 100$ Hz. The simulation timestep is 10^{-4} s and the trajectories are downsampled to $\Delta t = 10^{-3}$ s. The duration is $T = 300$ s. Such a long T is required to ensure that $\tilde{\mathcal{B}}_\Psi$ is close to $\mathcal{B}$. For N between 10 and 400 (by steps of 2, to ensure that N is even), we make a sliding window Δ_t such that $|\Delta_t| = N\Delta t$ and compute $\bar{x}_t$ (with the real $\mathcal{B}$), $\hat{x}_t$ and $\breve{x}_t$. The mean (over the 10 trajectories) root mean square error (rmse) between the true x_t trajectory and estimated one is computed for the three estimators and all N. The results are represented in figure 7.2. Given the smoothness of the results, 10 trajectories are indeed enough. We have even observed that 1 trajectory is enough: since it is 300 s long the rmse is already computed over about 300000 samples. The mean rmse for the three estimators have similar behaviors as they reach a minimum on the explored range (or almost for $\hat{x}_t$). The curves for $\bar{x}_t$ and $\breve{x}_t$ are almost identical, which means that using the true $\mathcal{B}$ or $\tilde{\mathcal{B}}_\Psi$ does not make much difference. The minimum mean rmse for $\bar{x}_t$ and $\breve{x}_t$ are much lower than that of $\hat{x}_t$. Figure 7.3 represents a portion of a trajectory of x_t, as well as $\bar{x}_t$, $\breve{x}_t$ and $\hat{x}_t$ at their respective minimizing N. It is clear that $\bar{x}_t$ and $\breve{x}_t$ are almost identical and are much better estimations than $\hat{x}_t$.

7.1.4   Optimal window size

We now want to explore the dependence of the optimal N on $\mathcal{A}$ and α for $\bar{x}_t$. This optimal N is denoted $N_{opt}(\mathcal{A}, \alpha, \Delta t)$. This dependence of N_{opt} on $\mathcal{A}$, α and Δt is found in [44], equation (32). It is derived analytically using an approximation for the expectation of the squared error between the true x_t and estimated one. Adapting the notations, this equation states:

$$\mathcal{N}_{opt} = 2 \left(\frac{12(\alpha + 1)}{\mathcal{A}\Delta t} \right)^{1/2}, \tag{7.26}$$

where $\mathcal{N}_{opt}$ stands for Δ^{opt} in the paper (half-window number of samples). Equation (7.26) gives us an idea of the form we should expect for the function N_{opt}. For $(\mathcal{A}, \alpha)$, we explore the space $\{0.1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}^2$, and for Δt, we explore $\{10^{-4}, 10^{-3}, 10^{-2}\}$. For fixed $(\mathcal{A}, \alpha)$, the trajectories are simulated with $\Delta t = 10^{-4}$ s and then downsampled to 10^{-3} and 10^{-2} s; their duration is 300 s. For fixed $(\mathcal{A}, \alpha, \Delta t)$, we vary N between 10 and 400, and select the N which minimizes the rmse between the true x_t and the estimation $\bar{x}_t$, where $\bar{x}_t$ is computed using a window Δ containing N samples. On a standard computer, the
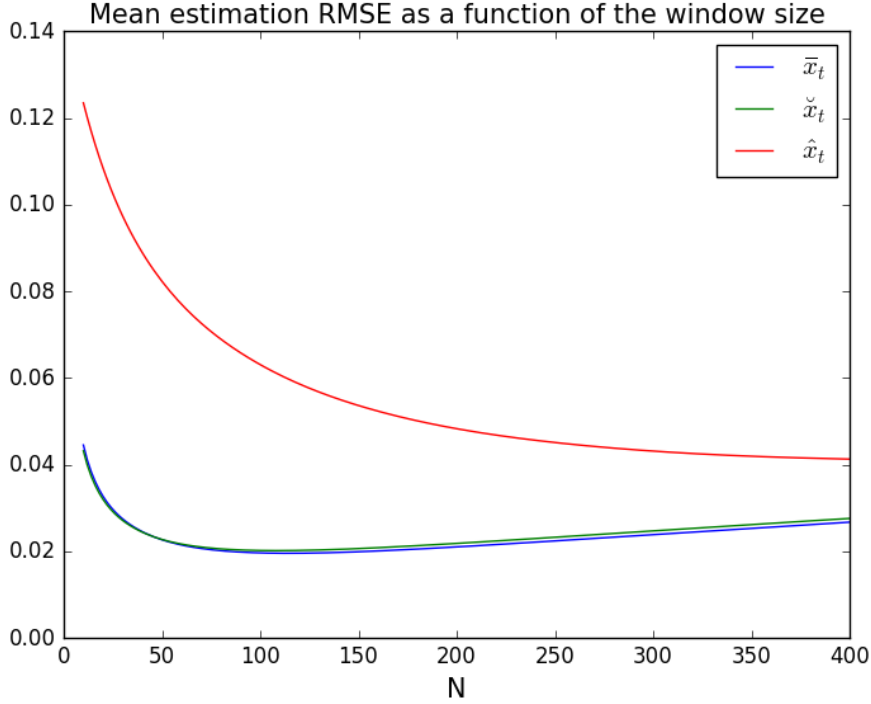
Figure 7.2: Mean rmse between the true x_t and its estimations as a function of window size.

simulations took one week. We represent sections of the obtained empirical $N_{opt}(\mathcal{A}, \alpha, \Delta t)$ for each of the three dimensions $(\mathcal{A}, \alpha, \Delta t)$ in figure 7.4. The value $\alpha = 0.1$ is voluntarily represented in a different figure due to its outlier-like behavior. Doing a simple ratio of the experimental N_{opt} and $\mathcal{N}_{opt}$ from equation (7.26), we find on average 0.32. Thus, we propose the theoretical expression for N_{opt}:

$$N_{opt} = 0.64 \left(\frac{12(\alpha + 1)}{\mathcal{A}\Delta t} \right)^{1/2}. \tag{7.27}$$

This theoretical N_{opt} is also represented in figure 7.4 for comparison. There is a very good agreement with the empirical N_{opt}, which encourages us to believe that if a rigorous analytical expression of N_{opt} exists, it should be proportional to $\mathcal{N}_{opt}$ and thus decreases as $\mathcal{A}^{-1/2}$, as $\Delta t^{-1/2}$, and increase as $(\alpha + 1)^{1/2}$.

Figure 7.5 represents the results for $\alpha = 0.1$. For all Δt, the empirical N_{opt} mainly follows the theoretical one, except for some value of $\mathcal{A}$ where it breaks out ($\mathcal{A} = 1$ Hz for $\Delta t = 0.01$ and 0.001 s, $\mathcal{A} = 5$ Hz for $\Delta t = 0.0001$ s). This observation would deserve more attention, but from now on, we simply consider that N_{opt} is given by equation (7.27).

7.1.5   Estimators for x_t and γ_t

Based on the results of section 7.1.3, $\breve{x}_t$ is the best estimator for x_t (since $\hat{x}_t$ is less performant and $\bar{x}_t$ not computable in practice). The corresponding estimator for γ_t is $\breve{\gamma}_t = \frac{\Psi_t}{\breve{x}_t^{1/2}}$. Finally, the retained estimators are:

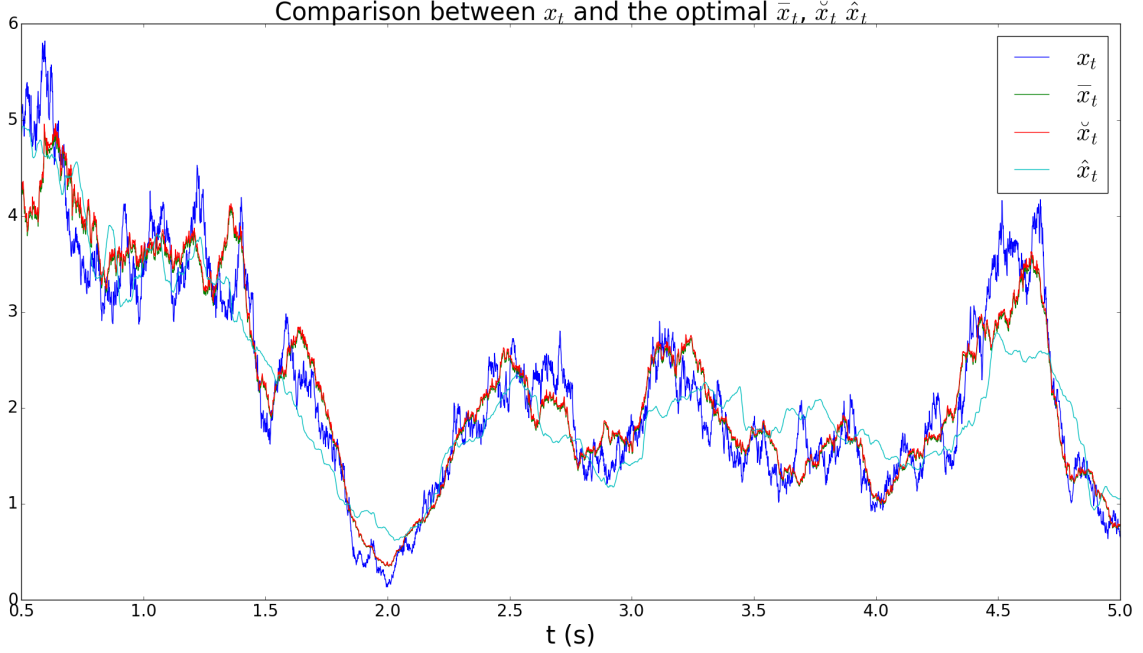$$\breve{x}_t = \frac{1}{\tilde{\mathcal{B}}_\Psi \Delta t N} \sum_{k \in \Delta_t} |\Delta_k \Psi|^2, \tag{7.28}$$

Figure 7.3: Example of a trajectory of x_t and its estimations $\bar{x}_t$, $\breve{x}_t$ and $\hat{x}_t$.

and

$$\breve{\gamma}_t = \frac{\Psi_t \left(\tilde{\mathcal{B}}_\Psi \Delta t N \right)^{1/2}}{\left(\sum_{k \in \Delta} |\Delta_k \Psi|^2 \right)^{1/2}}. \tag{7.29}$$

We also remind that:

$$\tilde{\mathcal{B}}_\Psi = \frac{1}{m \Delta t N} \sum_{i=1}^m \sum_{k \in \Delta_i} |\Delta_k \Psi|^2, \tag{7.30}$$

where N is the number of samples in window Δ_i (independent of i), and m the number of windows.

7.2 Sequential estimation of the clutter parameters, x_t and γ_t

In this section, we estimate the clutter parameters $\mathcal{A}$, $\mathcal{B}$ and α when x_t is not observed, as in real life. Therefore, we get rid of all the assumptions and finally tackle the full problem, without relying on unknown knowledge. Combining the results of the previous sections, as well as those of chapter 5, we can propose a very natural sequential estimation where the only data that we have is a discrete time series of $\Psi_t = R_t + iI_t$. The available data is therefore $\Psi = \{\Psi_k, k = 0, 1, 2, \ldots, n\}$, with $\Psi_k = \Psi_{t_k}$ and $t_k = k\frac{T}{n}$. Of course it is assumed that the sea clutter really follows Field's model and that therefore there are two objective but unobserved processes x_t and γ_t such that $\Psi_t = x_t^{1/2} \gamma_t$.

We now remind our notations. $\breve{x}_t$ and $\breve{\gamma}_t$ are the estimators of x_t and γ_t based on the increments of Ψ_t (equations (7.28) and (7.29)). $\tilde{\mathcal{A}}_{ML}$ and $\tilde{\mathcal{B}}_{ML}$ are the ML estimators for
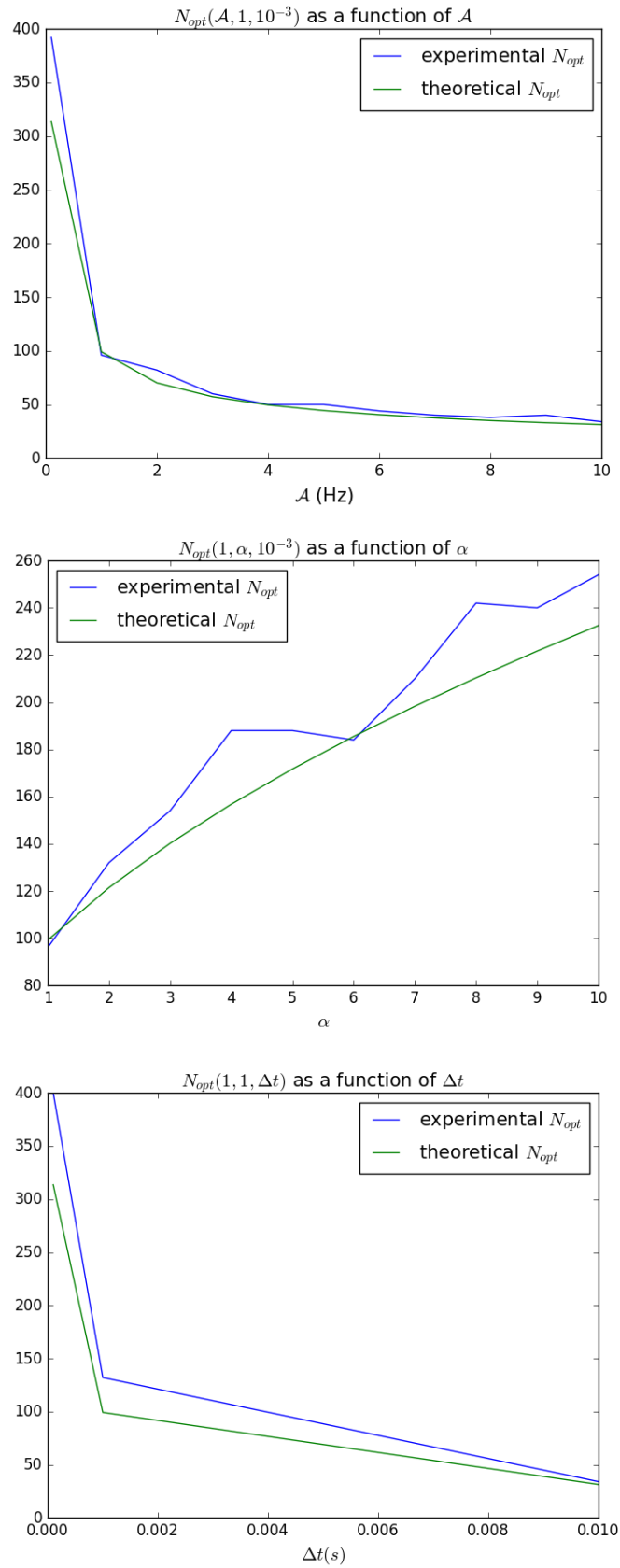
169

Figure 7.4: Empirical and theoretical N_{opt}. The theoretical N_{opt} is given by equation (7.27).
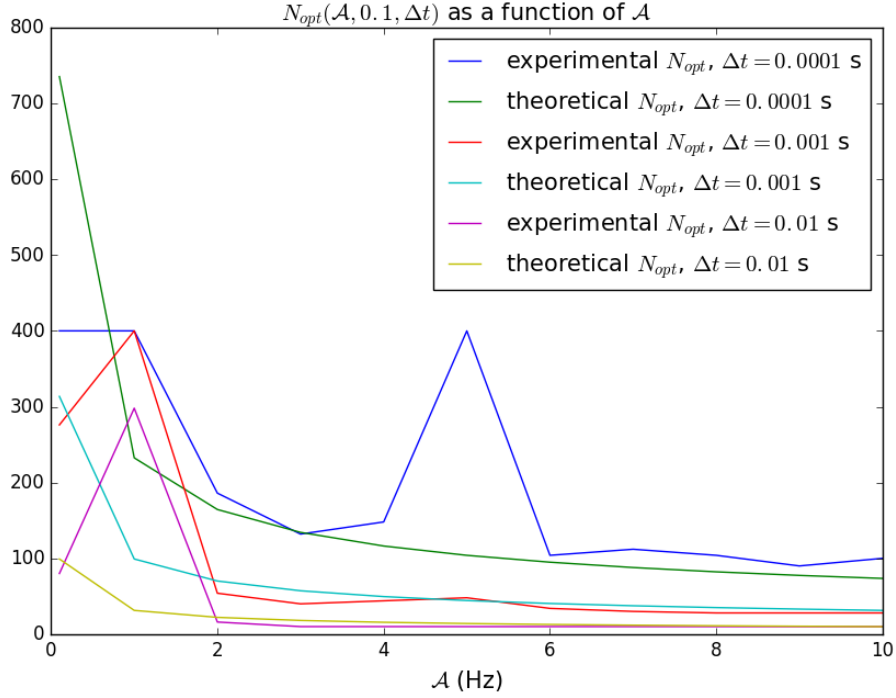
Figure 7.5: Empirical and theoretical N_{opt} for $\alpha = 0.1$. The theoretical N_{opt} is given by equation (7.27).

$\mathcal{A}$ and $\mathcal{B}$. $\tilde{\mathcal{B}}_\Psi$ is the estimator of $\mathcal{B}$ based on the increments of Ψ_t. Finally, $\tilde{\alpha}_x$ and $\tilde{\alpha}_z$ are respectively the estimators of α based on x_t and on $z_t = |\Psi_t|^2$ (see section 5.1.2).

7.2.1 Sequential estimation

Before presenting the sequential estimation, one preliminary step must be taken. Let us assume that we observe the data $\Psi = \{\Psi_k, k = 0, 1, 2, \ldots, n\}$. In echo with section 5.1.1, we acknowledge that *a priori*, Ψ (raw data) has no reason to be normalized, in the sense that:

$$C = \left[\frac{1}{n\Delta t} \sum_{k=0}^{n-1} |\Psi_k|^2 \Delta t \right]^{1/2} \tag{7.31}$$

is surely not equal or even close to 1. The first step consists in normalizing Ψ by C, to make sure that:

$$\left[\frac{1}{n\Delta t} \sum_{k=0}^{n-1} |\Psi_k|^2 \Delta t \right]^{1/2} = 1. \tag{7.32}$$

This normalization can be seen as a calibration step. It is possible also to work on a trajectory (for estimation purposes) shorter than the trajectory used for normalization. In all cases, it is assumed that for the whole duration of the normalization trajectory, the acquisition conditions do not change (constant mean power, *i.e.* stationarity). If the working trajectory is shorter (or just different) than the normalizing trajectory (still with stationarity), we will have:

$$\left[\frac{1}{n\Delta t} \sum_{k=0}^{n-1} |\Psi_k|^2 \Delta t \right]^{1/2} \approx 1, \tag{7.33}$$

where the approximation is due to the statistical effect.

We now propose the sequential estimation depicted in figure 7.6. We start from Ψ_t, measured by the radar (a normalized discrete time series of it, denote Ψ above). First, we compute $\tilde{\alpha}_z$ from $z_t = |\Psi_t|^2$ using equation (5.21). Second, we compute $\tilde{\mathcal{B}}_\Psi$ using equation (7.30). To do so, we use a sliding window which contains $N_{opt}(\mathcal{A}, \tilde{\alpha}_z, \Delta t)$ samples. The only unknown is $\mathcal{A}$, so we arbitrarily use 5 in N_{opt}. We then compute $\breve{x}_t$ using equation (7.28), of course using again $N_{opt}(5, \tilde{\alpha}_z, \Delta t)$ for the size of Δ. $\breve{\gamma}_t$ is immediately computed with equation (7.29).

At this stage, we have decoupled x_t and γ_t and we can now freshly estimate the clutter parameters. From $\breve{x}_t$, we can compute $\tilde{\mathcal{A}}_{ML}$ using equation (5.33). $\tilde{\alpha}_x$ can be computed using equation (5.13). From $\breve{\gamma}_t$, we can compute $\tilde{\mathcal{B}}_{ML}$ using equation (5.33).

Since we have now an estimation of $\mathcal{A}$, we can reloop by recomputing $\tilde{\mathcal{B}}_\Psi$ using $N_{opt}(\tilde{\mathcal{A}}_{ML}, \tilde{\alpha}_x, \Delta t)$. Of course it is possible to reloop a great many times.
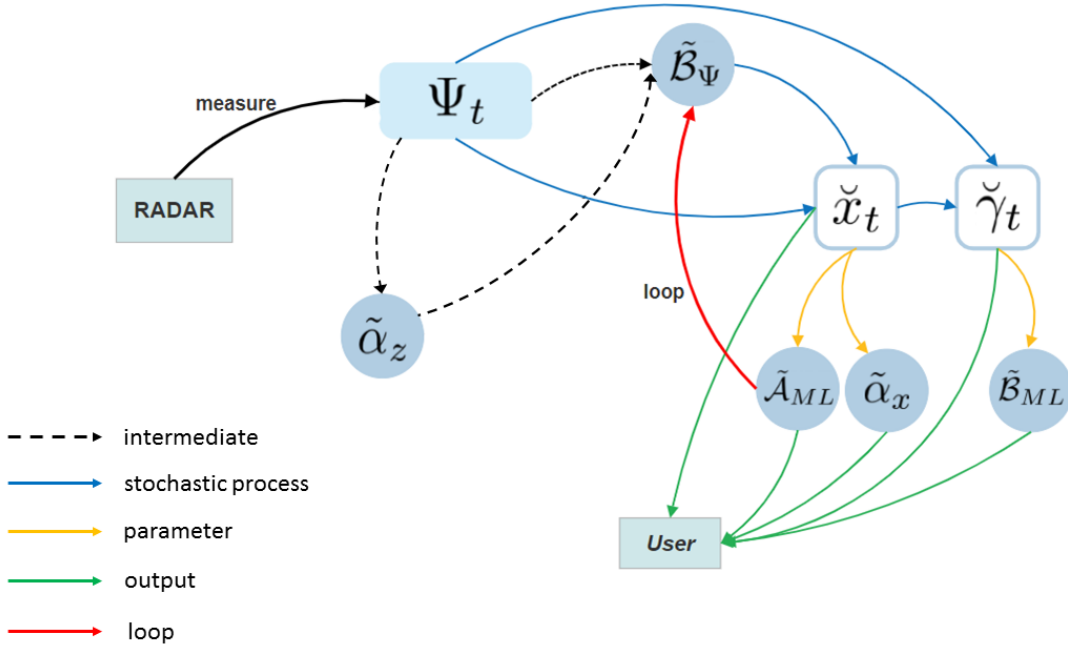
Figure 7.6: Sequence of estimators for estimating $\mathcal{A}$, $\mathcal{B}$, α, x_t and γ_t.

7.2.2 Numerical analysis of the proposed sequence of estimators

In this section, we want to analyse the sequence of estimators in figure 7.6 using numerical simulations. We are going to observe that once $\breve{x}_t$ has been computed, it is necessary to subsample it before estimating $\mathcal{A}$. We will also come across some counter intuitive results that we will clarify and discuss. For the analysis, we choose to reloop only once.

To assess the estimators for $\mathcal{A}$ and α, we have usually explored $\{0.1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ for both of them. For $\mathcal{B}$, we explored $\{10, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000\}$.

	$\tilde{\alpha}_z$	N_{opt}	$\tilde{\mathcal{B}}_\Psi$	rmse x_t	rmse γ_t	$\tilde{\mathcal{A}}_{ML}$	$\tilde{\alpha}_x$	$\tilde{\mathcal{B}}_{ML}$
1st round	0.998	44	97.675	0.227	0.121	0.527	0.991	94.773
2nd round	$n.a.$	136	97.666	0.191	0.148	0.064	1.053	93.517

Table 7.1: Results of the sequential estimation for the first trajectory without subsampling $\breve{x}_t$. $\mathcal{A} = 1$ Hz, $\alpha = 1$ and $\mathcal{B} = 100$ Hz. $n.a.$: not applicable.

Exploring the full parameter space

$$\{0.1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\} \times \{0.1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\} \times \{10, 100, 200, \dots, 1000\}$$

is not computationally realistic with a standard desktop computer. It represents 1000 parameters states, each of which requires to simulate 1000 trajectories which are 1200 s long (for estimating $\breve{\mathcal{B}}$). We prefer to set only $\mathcal{A} = 1$ Hz, $\alpha = 1$, $\mathcal{B} = 100$ Hz, $i.e.$ our defaults values. As indicated, we simulate 1000 trajectories with $T = 1200$ s and apply the sequence of estimators to each of those. The data is generated with a time step of $\hat{\Delta}t = 10^{-4}$ s and sampled at $\Delta t = 10^{-3}$ s before estimation.

Subsampling for estimating $\mathcal{A}$

Table 7.1 gathers the results for the first trajectory and two rounds (one reloop). rmse x_t is the root mean square error between x_t and $\breve{x}_t$, and rmse γ_t is the root mean square error between γ_t and $\breve{\gamma}_t$. For information, we have $\mathbb{E}[x_t^2] = 1 + \frac{1}{\alpha} = 2$, and $\mathbb{E}[|\gamma_t|^2] = 1$. We observe the first following facts. First $\tilde{\alpha}_z$ estimates very well α and $\tilde{\mathcal{B}}_\Psi$ estimates very well $\mathcal{B}$. Similarly, $\tilde{\mathcal{B}}_{ML}$ (ML estimator with Euler's approximation) and $\tilde{\alpha}_x$ perform well. Even though N_{opt} jumps from 44 to 136 between the two rounds, $\tilde{\mathcal{B}}_\Psi$, rmse x_t, rmse γ_t, $\tilde{\alpha}_x$ and $\tilde{\mathcal{B}}_{ML}$ are not significantly affected. However, $\mathcal{A}$ is very poorly estimated by $\tilde{\mathcal{A}}_{ML}$ (ML estimator with Euler's approximation). The second loop, is even worst than the first one. The first explain why the estimation (whatever the round), is an order of magnitude lower than the true value $\mathcal{A} = 1$ Hz.

The data is generated at times $t_i = i\Delta t$, for $i = 0, 1, \dots, n$, with $n = 1200.10^3$ here. $\breve{x}_t$ is computed at these same times, except that it is truncated of $\frac{N_{opt}}{2}$ samples at each sides. For each t_i, $\breve{x}_{t_i}$ is the average of N_{opt} increments $\Delta_k \Psi$ (equation (7.28)). Of course, two successive values, for example $\breve{x}_{t_i}$ and $\breve{x}_{t_{i+1}}$, have $N_{opt} - 1$ increments in common. Their values are therefore highly correlated and it is unlikely that their difference is described by the transition probability of x_t between t_i and t_{i+1}. Put simply, $\breve{x}_t$ is a smoothed version of x_t which does not capture its short time dynamics. Therefore, the ML estimator applied to the time series $\{\breve{x}_{t_l}, \breve{x}_{t_2}, \dots, \breve{x}_{t_m}\}$ observes that nothing happens over short times, suggesting a large correlation time, $i.e.$ a small $\mathcal{A}$. To solve this issue, we suggest to subsample $\breve{x}_t$ by a factor N_{opt}, to ensure that two successive values in the time series are computed using different increments $\Delta_k \Psi$. The new results, where subsampling is applied, are shown in table 7.2. We see that now $\tilde{A}$ is much closer to 1.

Analysis of the results after subsampling

Table 7.3 represents the estimators biases, computed over the 1000 trajectories. It also represents the average N_{opt} and average rmse x_t and rmse γ_t. These results are for $\mathcal{A} = 1$ Hz, $\alpha = 1$ and $\mathcal{B} = 1000$ Hz. The results do not add more to what was already in table 7.2.

	$\tilde{\alpha}_z$	N_{opt}	$\tilde{\mathcal{B}}_\Psi$	rmse x_t	rmse γ_t	$\tilde{\mathcal{A}}_{ML}$	$\tilde{\alpha}_x$	$\tilde{\mathcal{B}}_{ML}$
1st round	0.998	44	97.675	0.227	0.121	1.136	0.991	94.773
2nd round	*n.a.*	92	97.670	0.19	0.133	0.748	1.031	93.963

Table 7.2: Results of the sequential estimation for the first trajectory, of duration 1200 s. We have $\mathcal{A} = 1$ Hz, $\alpha = 1$ and $\mathcal{B} = 100$ Hz. $\breve{x}_t$ is subsampled by a factor N_{opt} before $\tilde{\mathcal{A}}_{ML}$ and $\tilde{\alpha}_x$ are computed. *n.a.* : not applicable.

	$b(\tilde{\alpha}_z)$	$\langle N_{opt} \rangle$	$b(\tilde{\mathcal{B}}_\Psi)$	$\langle$rmse $x_t\rangle$	$\langle$rmse $\gamma_t\rangle$	$b(\tilde{\mathcal{A}}_{ML})$	$b(\tilde{\alpha}_x)$	$b(\tilde{\mathcal{B}}_{ML})$
1st round	0.007	43.472	−1.734	0.236	0.123	0.138	−0.014	−4.695
2nd round	*n.a.*	91.664	−1.738	0.201	0.135	−0.276	0.025	−5.481

Table 7.3: Estimation bias of the estimators, average N_{opt}, rmse x_t and rmse γ_t over 1000 trajectories of duration 1200 s, with $\mathcal{A} = 1$ Hz, $\alpha = 1$ and $\mathcal{B} = 100$ Hz. *n.a.* : not applicable.

The same color code can therefore be adopted. Table 7.4 represents the estimation standard deviations. They do not seem to be able to discriminate between the estimators and rounds which is why we are more focused on table 7.3.

In our configuration, it seems that:

- the first round is better than the second one, especially for $\tilde{\mathcal{A}}_{ML}$

- $\tilde{\alpha}_z$ and $\tilde{\mathcal{B}}_\Psi$ are better than $\tilde{\alpha}_x$ and $\tilde{\mathcal{B}}_{ML}$.

These results are unexpected and force us to go a bit further in our analysis. We decide to apply the sequential estimation to shortened versions of the 1000 trajectories we simulated. We apply it to $T = 30$ s and $T = 300$ s ($\mathcal{A} = 1$ Hz, $\alpha = 1$ and $\mathcal{B} = 100$ Hz). The estimation bias and standard deviation for $\tilde{\mathcal{A}}_{ML}$, $\tilde{\alpha}_x$ and $\tilde{\alpha}_z$ are gathered in table 7.5. The estimation standard deviations decrease as T increases (expected). However, the estimation bias of $\tilde{\mathcal{A}}_{ML}$ increases (unexpected).

Explanation of the results for $\tilde{\mathcal{A}}_{ML}$

We now explain why the estimation of $\mathcal{A}$ seems to deteriorate when we loop (second round). We have generated time series of the form $\{x_{t_0}, x_{t_1}, \ldots, x_{t_n}\}$ with $\mathcal{A} = 1$ Hz and $\alpha = 1$. If we were estimating $\mathcal{A}$ using these time series, we would expect to find $\tilde{\mathcal{A}}_{ML}$ close to 1. However, we need to work on $\{\Psi_{t_0}, \Psi_{t_1}, \ldots, \Psi_{t_n}\}$ and have to compute $\{\breve{x}_{t_l}, \breve{x}_{t_2}, \ldots, \breve{x}_{t_m}\}$, *i.e.* to estimate x_t, before estimating $\mathcal{A}$. Though $\mathcal{A} = 1$ Hz explains well $\{x_{t_0}, x_{t_1}, \ldots, x_{t_n}\}$, we claim that the biased $\tilde{\mathcal{A}}_{ML}$ (which is on average $1 - 0.276 = 0.724$ Hz) simply explains better $\{\breve{x}_{t_l}, \breve{x}_{t_2}, \ldots, \breve{x}_{t_m}\}$, and that this is the reason why it was obtained in the first place.

	$\sigma(\tilde{\alpha}_z)$	$\sigma(\tilde{\mathcal{B}}_\Psi)$	$\sigma(\tilde{\mathcal{A}}_{ML})$	$\sigma(\tilde{\alpha}_x)$	$\sigma(\tilde{\mathcal{B}}_{ML})$
1st round	0.063	4	0.057	0.057	3.728
2nd round	*n.a.*	3.999	0.049	0.06	3.701

Table 7.4: Estimation standard deviation of the estimators over 1000 trajectories of duration 1200 s, with $\mathcal{A} = 1$ Hz, $\alpha = 1$ and $\mathcal{B} = 100$ Hz. *n.a.* : not applicable.

	$\sigma(\tilde{\alpha}_z)$	$\sigma(\tilde{\mathcal{A}}_{ML})$	$\sigma(\tilde{\alpha}_x)$	$b(\tilde{\alpha}_z)$	$b(\tilde{\mathcal{A}}_{ML})$	$b(\tilde{\alpha}_x)$
$T = 30$ s, 1st round	0.389	0.313	0.328	0.253	0.32	0.196
$T = 30$ s, 2nd round	*n.a.*	0.28	0.343	*n.a.*	-0.114	0.242
$T = 300$ s, 1st round	0.12	0.109	0.11	0.029	0.155	0.005
$T = 300$ s, 2nd round	*n.a.*	0.093	0.116	*n.a.*	-0.262	0.045
$T = 1200$ s, 1st round	0.063	0.057	0.057	0.007	0.138	-0.014
$T = 1200$ s, 2nd round	*n.a.*	0.049	0.06	*n.a.*	-0.276	0.025

Table 7.5: Estimation bias and standard deviation of $\tilde{\mathcal{A}}_{ML}$, $\tilde{\alpha}_x$ and $\tilde{\alpha}_z$ for two rounds and $T = 30$ s, $T = 300s$ and $T = 1200$ s. The results are for 1000 trajectories with $\mathcal{A} = 1$ Hz, $\alpha = 1$ and $\mathcal{B} = 100$ Hz. *n.a.* : not applicable.

	$\breve{x}_t, (\mathcal{A}, \alpha)$	$\breve{x}_t, (\tilde{\mathcal{A}}_{ML}, \tilde{\alpha}_x)$	$\breve{x}_t, (\tilde{\mathcal{A}}_{ML}, \tilde{\alpha}_z)$	$x_t, (\mathcal{A}, \alpha)$
$T = 30$ s, 1st round	2.192	2.162	2.174	1.97
$T = 30$ s, 2nd round	2.359	2.506	2.504	*n.a.*
$T = 300$ s, 1st round	2.189	2.154	2.16	1.976
$T = 300$ s, 2nd round	2.378	2.525	2.518	*n.a.*
$T = 1200$ s, 1st round	2.189	2.153	2.158	1.976
$T = 1200$ s, 2nd round	2.379	2.526	2.519	*n.a.*

Table 7.6: Average log transition probability of $\breve{x}_t$ for $(\mathcal{A}, \alpha)$, $(\tilde{\mathcal{A}}_{ML}, \tilde{\alpha}_x)$ and $(\tilde{\mathcal{A}}_{ML}, \tilde{\alpha}_z)$, and of x_t for $(\mathcal{A}, \alpha)$. *n.a.* : not applicable.

To show that, we define the average logarithm of the transition probabilities (or likelihoods) as:

$$\bar{l}(\breve{x}, \mathcal{A}, \alpha) = \frac{1}{m - l} \sum_{k=l}^{m-1} \ln p_{\mathcal{A}, \alpha}(\breve{x}_{t_{k+1}} \mid \breve{x}_{t_k}). \tag{7.34}$$

We choose $\bar{l}(\breve{x}, \mathcal{A}, \alpha)$ as a measure of how well $\mathcal{A}$ and α explain the transitions observed in $\breve{x}$. We compute it for each of the trajectories, and then average over the trajectories. We compare the couples $(\mathcal{A}, \alpha)$, $(\tilde{\mathcal{A}}_{ML}, \tilde{\alpha}_x)$ and $(\tilde{\mathcal{A}}_{ML}, \tilde{\alpha}_z)$ for $T = 30, 300$ and 1200 s.

The results are in table 7.6. They confirm our claim, since we now observe that the second rounds are significantly better than the first ones, and that $\tilde{\mathcal{A}}_{ML}$ is actually better than $\mathcal{A}$ to explain the transitions of $\breve{x}$. Using $\tilde{\alpha}_x$ or $\tilde{\alpha}_z$ is unimportant. By definition, the average log transition probability has low sensibility to T. We also computed the average log transition probability of x_t with the true $(\mathcal{A}, \alpha)$. Though it is of the same order of magnitude as the results for $\breve{x}_t$, it is a bit lower.

Discussion

We have seen that direct application of the estimation sequence depicted in figure 7.6 is able to estimate very well α and $\mathcal{B}$ already in the first round. However, the order of magnitude of $\tilde{\mathcal{A}}_{ML}$ was too low, which was solved by subsampling. Surprisingly, the second round seemed to give poorer results than the first one for $\tilde{\mathcal{A}}_{ML}$. Nonetheless, the average log transition probability increases, which shows progression in explaining the observed transitions. We think that there is an extremely interesting situation here. As explained before, $\tilde{\mathcal{A}}_{ML}$, whatever its value, is simply the best to explain $\breve{x}_t$ (even better than $\mathcal{A}$). Of course it is not better than $\mathcal{A}$ to explain the true RCS x_t. We think that subsampling still makes sense since $\breve{x}_t$ cannot reproduce the short-time dynamics of x_t.

We could continue tuning the estimation sequence to make $\tilde{\mathcal{A}}_{ML}$ closer to 1. We are not entirely convinced that it is useful: our ultimate goal is not to estimate $\mathcal{A}$. Our goal is to get a consistent and precise description of the sea clutter, to develop signal processing tools (see for example section 7.4 for target detection). We will never work on x_t since it is not observable, we will work on $\breve{x}_t$. Even if we tried to make it as close as possible to x_t by tuning N (section 7.1.4), the hypothesized objective truth is eventually out of reach. In practice, the objective (x_t, γ_t) decomposition of the sea clutter steps down to make place to the subjective decomposition $(\breve{x}_t, \breve{\gamma}_t)$.

We noticed that $\tilde{\alpha}_z$ and $\tilde{\mathcal{B}}_\Psi$ seemed better than $\tilde{\alpha}_x$ and $\tilde{\mathcal{B}}_{ML}$. We think that the estimators for α are equivalent. For $\mathcal{B}$, we can develop the same line of argument as for $\mathcal{A}$. Though seemingly $\tilde{\mathcal{B}}_\Psi$ outperforms $\tilde{\mathcal{B}}_{ML}$, $\tilde{\mathcal{B}}_{ML}$ better describes the transition of $\breve{\gamma}_t$ and should therefore be preferred.

More effort should be done to properly understand the role of the average log transition probability in relation with information theory and the concept of entropy. Why is it lower for x_t and $(\mathcal{A}, \alpha)$? Does that make sense? Also, we want to remind that though $\tilde{\mathcal{A}}_{ML}$ is the best parameter to explain $\breve{x}_t$, this is true under the hypothesis that $\breve{x}_t$ follows the SDE of x_t:

$$\mathrm{d}x_t = \mathcal{A}(1 - x_t)\mathrm{d}t + \left(2\frac{\mathcal{A}}{\alpha}x_t\right)^{\frac{1}{2}}\mathrm{d}W_t^{(x)}. \tag{7.35}$$

The question of whether it is indeed a good description of $\breve{x}_t$ is a different matter (model testing) which should be worked out, again presumably with information theory. The question also holds for $\breve{\gamma}_t$ and $\tilde{\mathcal{B}}_{ML}$. Because of that, we shall recognize that our work on the estimation of x_t, γ_t, $\mathcal{A}$, $\mathcal{B}$ and α is incomplete despite that the factorization is perfect by construction:

$$\Psi_t = \breve{x}_t^{1/2}\breve{\gamma}_t. \tag{7.36}$$

7.3 Bayesian estimation of the clutter parameters

We now succinctly introduce the topic of bayesian estimation of the clutter parameters.

From chapter 5 until now, we have derived estimators for the clutter parameters, target parameters, and for x_t and γ_t. They were all expressed in a similar way, as an explicit function of the data (discrete time series). Since the data is random, the estimator is a random variable. We always assessed our capacity to estimate well from the data by assessing the "performance" of the estimator. This performance was quantified either by the estimation bias and standard deviation, or by the root mean square error after debiasing. Overall, the choice of estimator and the way to measure its performance do not change the general situation: estimation has been entirely conceptualized from a frequentist point of view so far. Let θ be some parameter to be estimated (not the phase of the complex reflectivity), and x some data which contains information about θ. We have proposed formula of the type:

$$\tilde{\theta} = f(x), \tag{7.37}$$

where $\tilde{\theta}$ is an estimator of θ and f is some function. We assumed that there is a true θ out there, and we approximate it by a random variable built from our random data. The bayesian point of view, whose core is Bayes formula, adopts a different attitude. Though

it is highschool level mathematics, there is deep meaning when quoted in a data analysis context. Bayes formula asserts:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}. \tag{7.38}$$

$p(\theta|x)$ is the *posterior* distribution of θ, $p(x|\theta)$ is the likelihood, $p(\theta)$ is the *prior* distribution of θ, and $p(x)$ is the *marginal likelihood*. The main innovation is that we now speak of distributions for θ itself, not some estimators of it. These distributions refer to the knowledge that we have on θ, such that the bayesian perspective is more subjective than the frequentist one (if not completely subjective, as opposed to an objective frequentist point of view). $p(x \mid \theta)$ is the likelihood and is precisely the function we maximized for maximum likelihood estimation in chapters 5 and chapter 6. $p(x)$ is usually ignored because it is a normalizing constant with respect to θ (does not depend on θ). We have:

$$p(x) = \int p(x|\theta)p(\theta)\mathrm{d}\theta. \tag{7.39}$$

$p(\theta)$ is the knowledge on θ before observing the data x, and $p(\theta|x)$ is the knowledge on θ after x has been observed and taken into account. It is expected that $p(\theta|x)$ is narrower than $p(\theta)$, in which case Bayes formula really describes *learning* (see figure 7.7). The bayesian point of view is now widely used expecially in topics without "big" data (*e.g.* paleontology). Interestingly enough, bayesianism is at the core of a recent interpretation of quantum mechanics: QBism (*i.e.* quantum bayesianism, see [138] and the articles of Christopher A. Fuchs).

We can use equation (7.38) to estimate the clutter parameters from x_t and γ_t. For example, let $\theta = (\mathcal{A}, \alpha)$ and $x = \{x_{t_0}, x_{t_1}, \ldots, x_{t_n}\}$. Bayes formula becomes:

$$p(\mathcal{A}, \alpha|x) \propto p(x|\mathcal{A}, \alpha)p(\mathcal{A})p(\alpha), \tag{7.40}$$

assuming independence of the priors on $\mathcal{A}$ and α. Based on our knowledge of the range of possible values for $\mathcal{A}$ and α, we would suggest uniform prior for $\mathcal{A}$ over $[0.1, 10]$, *i.e.* $p(\mathcal{A}) \sim U(0.1, 10)$, and a gamma distributed prior for α with shape parameter 3 and rate $1/2$: $p(\alpha) \sim \Gamma(3, 1/2)$. We also know that the likelihood of x can be written:

$$p(x|\mathcal{A}, \alpha) = p_{\mathcal{A}, \alpha}(x_{t_0}) \prod_{k=1}^{n} p_{\mathcal{A}, \alpha}(x_{t_k}|x_{t_{k-1}}), \tag{7.41}$$

where $p_{\mathcal{A}, \alpha}(x_{t_0})$ is the likelihood of the first value of the time series (asymptotic distribution) and the $p_{\mathcal{A}, \alpha}(x_{t_k}|x_{t_{k-1}})$ are transition probabilities. Equation (7.41) is a "random" likelihood, explaining the slightly different notation from the previous chapters. However, it is really the same likelihood as in chapter 5. Equation (7.41) can be rewritten:

$$p(x|\mathcal{A}, \alpha) = \frac{\alpha^{\alpha} x_{t_0}^{\alpha-1} e^{-\alpha x_{t_0}}}{\Gamma(\alpha)} \prod_{k=1}^{n} \frac{\sqrt{\alpha}}{\sqrt{4\pi x_{t_{k-1}} \mathcal{A}\Delta t}} e^{-\frac{\alpha\left(x_{t_k} - \mathcal{A}\Delta t - (1-\mathcal{A}\Delta t)x_{t_{k-1}}\right)^2}{4\mathcal{A}\Delta t x_{t_{k-1}}}}. \tag{7.42}$$

In equation (7.42), we have used Euler-Maruyama's approximation for the transition probabilities (see chapter 5). It holds under the assumption that $\Delta t = t_k - t_{k-1}$ is small, but if this is not so it is always possible to use the exact transition probabilities derived in chapter
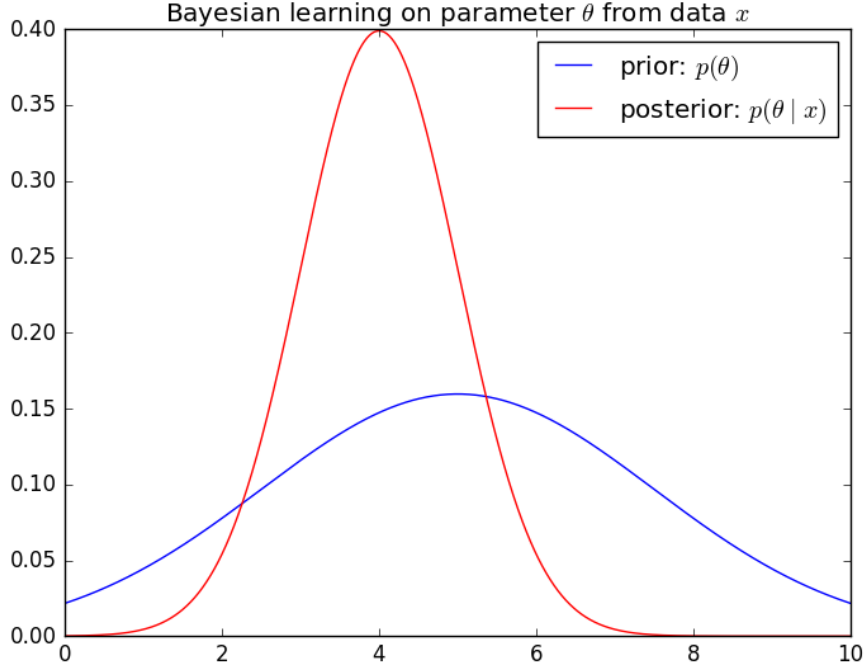
Figure 7.7: Illustrative example of Bayes formula. The prior distribution (blue) is large, suggesting poor knowledge on parameter θ. Once the data x has been observed, the knowledge on θ is summarized in the posterior distribution (red), which is a shifted and compressed version of the prior here. The relation between the prior and posterior is often more complicated in reality, but it is expected that the posterior has lower variance, suggesting improved knowledge.

4. We see that our efforts carried out in the frequentist world are useful for the bayesian one. In theory, we have all the elements to compute $p(\mathcal{A}, \alpha | x)$. In practice, the product of the prior and of the likelihood has (most of the time) unknown properties. However, we could draw samples from it and get histograms of $p(\mathcal{A}, \alpha | x)$, $p(\mathcal{A} | x) = \int p(\mathcal{A}, \alpha | x) \mathrm{d}\alpha$ and $p(\alpha | x) = \int p(\mathcal{A}, \alpha | x) \mathrm{d}\mathcal{A}$ using Metropolis–Hastings Markov chain Monte Carlo (MCMC).

Bayesian estimation can also be done for $\mathcal{B}$ using a time series of $\gamma_t^{(R)}$ for example: $\gamma = \{\gamma_{t_0}^{(R)}, \gamma_{t_1}^{(R)}, \ldots, \gamma_{t_n}^{(R)}\}$. We get:

$$p(\mathcal{B} | \gamma) \propto p(\gamma | \mathcal{B}) p(\mathcal{B}). \tag{7.43}$$

We would simply choose a uniform prior over $[10, 1000]$: $p(\mathcal{B}) \sim U(10, 1000)$. The likelihood can be expressed as:

$$p(\gamma | \mathcal{B}) = p_{\mathcal{B}}(\gamma_{t_0}^{(R)}) \prod_{k=1}^{n} p_{\mathcal{B}}(\gamma_{t_k}^{(R)} | \gamma_{t_{k-1}}^{(R)}). \tag{7.44}$$

Using Euler-Maruyama's approximation (see chapter 5), the likelihood becomes:

$$p(\gamma | \mathcal{B}) = \frac{1}{\sqrt{\pi}} e^{-\gamma_{t_0}^{(R)\,2}} \prod_{k=1}^{n} \frac{1}{\sqrt{\pi \mathcal{B} \Delta t}} e^{-\frac{\left(\gamma_{t_k}^{(R)} - \gamma_{t_{k-1}}^{(R)}(1 - \mathcal{B}\Delta t/2)\right)^2}{\mathcal{B}\Delta t}}. \tag{7.45}$$

178

MCMC could again be used to draw samples from $p(\mathcal{B}|\gamma)$.

It would be interesting to apply the bayesian estimation we just described for $\mathcal{A}$ and $\mathcal{B}$ (*e.g.* assuming α is known), and to use MCMC to compute the *a posteriori* distributions of the parameters. The means and standard deviations of these distributions could then be compared to the results of chapter 5 (estimation bias and standard deviation of ML estimators).

7.4   Target detection

One of the most important topics related to sea clutter is target detection for maritime surveillance. We have briefly mentionned target detection in chapter 6 section 6.4, where the target was modelled by a complex constant Ψ_c and estimated by maximum likelihood. Instead of talking about a target, we shall sometimes refer to anomaly detection, which is more neutral. A surfaced whale should not deserve to be treated as a target (see [3]), though recent news suggest it could sadly be the case [4].

We wish now to introduce the problematic of target detection in the framework of Field's model. We assume that the parameters $\mathcal{A}$, $\mathcal{B}$ and α are known, or well estimated. In section 7.4.1, we present a method based on spot volatility, which is based on observable quantities: Ψ_t, $z_t = |\Psi_t|^2$ and θ_t (phase of the complex reflectivity). No parameter estimation is required. In section 7.4.2, we propose to use the likelihood of the time series to detect anomalies.

7.4.1   Detection based on spot volatility

We propose the following method based on spot volatility (see below) to detect anomalies in sea clutter.

In Field's model, the complex reflectivity Ψ_t (sea clutter) solves:

$$
\begin{aligned}
\mathrm{d}\Psi_t &= -\frac{1}{2}\mathcal{B}\Psi_t\mathrm{d}t + \frac{\mathcal{B}^{1/2}}{\sqrt{2}}x_t^{1/2}\mathrm{d}W_t^{(R)} + i\frac{\mathcal{B}^{1/2}}{\sqrt{2}}x_t^{1/2}\mathrm{d}W_t^{(I)} \\
&+ \mathcal{A}\Psi_t\left(\frac{2(\alpha - x_t) - 1}{4x_t}\right)\mathrm{d}t + \Psi_t\left(\frac{\mathcal{A}}{2x_t}\right)^{1/2}\mathrm{d}W_t^{(x)},
\end{aligned} \tag{7.46}
$$

where x_t has replaced αx_t (equation (8.12) in [48]). From equation (7.46), we can show that:

$$
\mathrm{d}\langle\Psi, \Psi^*\rangle_t = \left(\frac{\mathcal{A}z_t}{2x_t} + \mathcal{B}x_t\right)\mathrm{d}t. \tag{7.47}
$$

Ψ_t^* is the complex conjugate of Ψ_t. Equation (7.47) is equation (8.36) in [48].

Using the SDEs for x_t and γ_t and Itô calculus, we can show that the intensity z_t solves:

$$
\mathrm{d}z_t = \left[\mathcal{B}(x_t - z_t) + \frac{\mathcal{A}z_t(\alpha - x_t)}{x_t}\right]\mathrm{d}t + \left(2\mathcal{B}x_tz_t + \frac{2\mathcal{A}z_t^2}{x_t}\right)^{1/2}\mathrm{d}W_t^{(z)}, \tag{7.48}
$$

which is equation (8.18) in [48]. $W_t^{(z)}$ is an undefined brownian motion which depends on $W_t^{(x)}$, $W_t^{(R)}$ and $W_t^{(I)}$.

179

From equation (7.48), we get:

$$d\langle z \rangle_t = \left(2\mathcal{B}x_t z_t + \frac{2\mathcal{A}z_t^2}{x_t} \right) dt = \sigma_t^{(z)} dt, \tag{7.49}$$

with $\sigma_t^{(z)} = \left(2\mathcal{B}x_t z_t + \frac{2\mathcal{A}z_t^2}{x_t} \right)$. $\sigma_t^{(z)}$ is called the *spot volatility*, and it can be estimated from a discrete time series of z_t (see [8]). Its estimation is more delicate than that of the integrated volatility, and it has slower convergence rate.

Finally, if we write $\Psi_t = r_t e^{i\theta_t}$, Field shows that the SDE for the phase is:

$$d\theta_t = \left(\frac{\mathcal{B}x_t}{2z_t} \right)^{1/2} dW_t^{(\theta)}, \tag{7.50}$$

where $W_t^{(\theta)}$ is some brownian motion. This is equation (8.29) in [48]. From equation (7.50):

$$d\langle \theta \rangle_t = \frac{\mathcal{B}x_t}{2z_t} dt = \sigma_t^{(\theta)} dt, \tag{7.51}$$

where $\sigma_t^{(\theta)} = \frac{\mathcal{B}x_t}{2z_t}$ is the spot volatility of θ_t. Since θ_t is actually observed, its spot volatility can be estimated.

From equations (7.47), (7.49) and (7.51), we get:

$$d\langle \Psi, \Psi^* \rangle_t = \left(\frac{\sigma_t^{(z)}}{4z_t} + z_t \sigma_t^{(\theta)} \right) dt. \tag{7.52}$$

In integral notations, we obtain:

$$\langle \Psi, \Psi^* \rangle_T = \int_0^T \left(\frac{\sigma_t^{(z)}}{4z_t} + z_t \sigma_t^{(\theta)} \right) dt. \tag{7.53}$$

We now assume that we observe a discrete time series $\{z_k, k = 1, 2, \ldots, n\}$ and $\{\theta_k, k = 1, 2, \ldots, n\}$ at times $t_k = k\frac{T}{n} = k\Delta t$. From these, it is also possible to estimate the spot volatilities. We note respectively $\tilde{\sigma}_k^{(z)}$ and $\tilde{\sigma}_k^{(\theta)}$ the estimators of $\sigma_{t_k}^{(z)}$ and $\sigma_{t_k}^{(\theta)}$. By definition, we have:

$$\langle \Psi, \Psi^* \rangle_T = \lim_{n \to +\infty} \sum_{k=1}^{n} |\Psi_{t_k} - \Psi_{t_{k-1}}|^2, \tag{7.54}$$

and using a simple Riemann sum to approximate the integral in the right hand side of (7.53), we get:

$$\sum_{k=1}^{n} |\Psi_{t_k} - \Psi_{t_{k-1}}|^2 \approx \sum_{k=1}^{n} \left(\frac{\tilde{\sigma}_k^{(z)}}{4z_k} + z_k \tilde{\sigma}_k^{(\theta)} \right) \Delta t. \tag{7.55}$$

Equation (7.55) is the basis for hypothesis testing. Under H_0 (sea clutter only, no target), we expect that:

$$\left| \sum_{k=1}^{n} |\Psi_{t_k} - \Psi_{t_{k-1}}|^2 - \sum_{k=1}^{n} \left(\frac{\tilde{\sigma}_k^{(z)}}{4z_k} + z_k \tilde{\sigma}_k^{(\theta)} \right) \Delta t \right| = 0. \tag{7.56}$$

Of course, it cannot hold perfectly but it should be close to zero. In practice, we should set a threshold for the left hand side of equation (7.56) and reject H_0, *i.e.* detect an anomaly, if the threshold is exceeded.

This method has the huge advantage of not requiring any hypothesis on the potentially present target, and of being applicable to observable quantities only. Its drawback is that it requires spot volatility estimation, which works only for high sampling frequencies.

7.4.2 Detection based on likelihood

First approach: parameter estimation

The first approach we propose is to rely on parameter estimation. In chapter 6 and in section 7.2, the target is represented by a constant Ψ_c. We considered two models for the total reflectivity: HK scattering ($\Psi_t = x_t^{1/2}\gamma_t + \Psi_c$) and GK scattering ($\Psi_t = x_t^{1/2}\gamma_t + \eta x_t \Psi_c$). In both cases, we assumed that:

1. the target Ψ_c is constant,

2. discrete time series of x_t, R_t, I_t are observed,

and then estimated Ψ_c either by ML (chapter 6), ergodicity (chapter 6), or by volatility (appendix A.2). The constant target approximation was reasonable since we considered 1 s long trajectories. However, we know that x_t is not observed. We now consider a time-dependent target $\Psi_c(t)$. We have:

$$\Psi_t = x_t^{1/2}\gamma_t + \Psi_c(t) \qquad (HK\ scattering) \tag{7.57}$$

$$\Psi_t = x_t^{1/2}\gamma_t + x_t\Psi_c(t) \quad (GK\ scattering). \tag{7.58}$$

For illustrative purposes only, we choose the following exemple of real (no imaginary part) $\Psi_c(t)$:

$$\begin{cases} \Psi_c(t) = \Psi_c^{(R)} e^{-\frac{1}{r^2-(t-t_c)^2}} & if\ r^2-(t-t_c)^2 > 0 \\ \Psi_c(t) = 0 & otherwise. \end{cases} \tag{7.59}$$

This function is an example of C^∞ but non analytic function. It takes non zero values for $r^2-(t-t_c)^2 \geq 0$, *i.e.* for $|t-t_c| < r$. t_c is therefore the central time of the target, and its width (lifetime) is $2r$. Of course, the way a target really appears in terms of reflectivity is a different topic which deserves considerable attention on its own. If x_t was observed, we could estimate $\Psi_c(t)$ using a sliding window. The estimator $\breve{x}_t$ (or $\bar{x}_t$ if $\mathcal{B}$ has already been estimated) for x_t defined in section 7.1 can be extended when a target of the form (7.59) is present, but in HK scattering only. Indeed, the reflectivity increments are then:

$$\Delta_k\Psi = \Psi_{t_k} - \Psi_{t_{k-1}} = x_{t_k}^{1/2}\gamma_{t_k} - x_{t_{k-1}}^{1/2}\gamma_{t_{k-1}} + \Psi_c(t_k) - \Psi_c(t_{k-1}), \tag{7.60}$$

and if $\Psi_c(t)$ is smooth, we may neglect $\Psi_c(t_k) - \Psi_c(t_{k-1})$ and resume the precedent derivation to arrive at the estimator $\bar{x}_t$ or $\breve{x}_t$. We show a simple numerical example with a 30 s long trajectory, including a target of width $2r = 5$ s, central time $t_c = 15$ s, and $\Psi_c^{(R)} = 1$, in figure 7.8. The top represents the target model. The middle is the time series of the real part of the reflectivity including the target. The bottom represents the estimated target with a sliding window of 1 s, using ML and ergodicity. They are respectively denoted $\tilde{\Psi}_{c,ML}(t)$ and $\tilde{\Psi}_{c,e}(t)$. The volatility-based estimation is not represented since for HK scattering (additive target) it is outperformed by the other two estimators. We use a 0.5 s wide sliding window for estimation, and the target signal is very well retrieved (bottom) even though it was almost hidden in sea clutter (middle). We have used the true x_t for simplicity. In practice, it is required to estimate x_t first.

We think that detection schemes could be developped based on $\tilde{\Psi}_{c,ML}(t)$ or $\tilde{\Psi}_{c,e}(t)$, knowing that it should normally be around zero under the hypothesis H_0 that there is sea clutter only. A drawback of this approach is that it assumes that the presence of an anomaly appears as an additive term in the reflectivity.
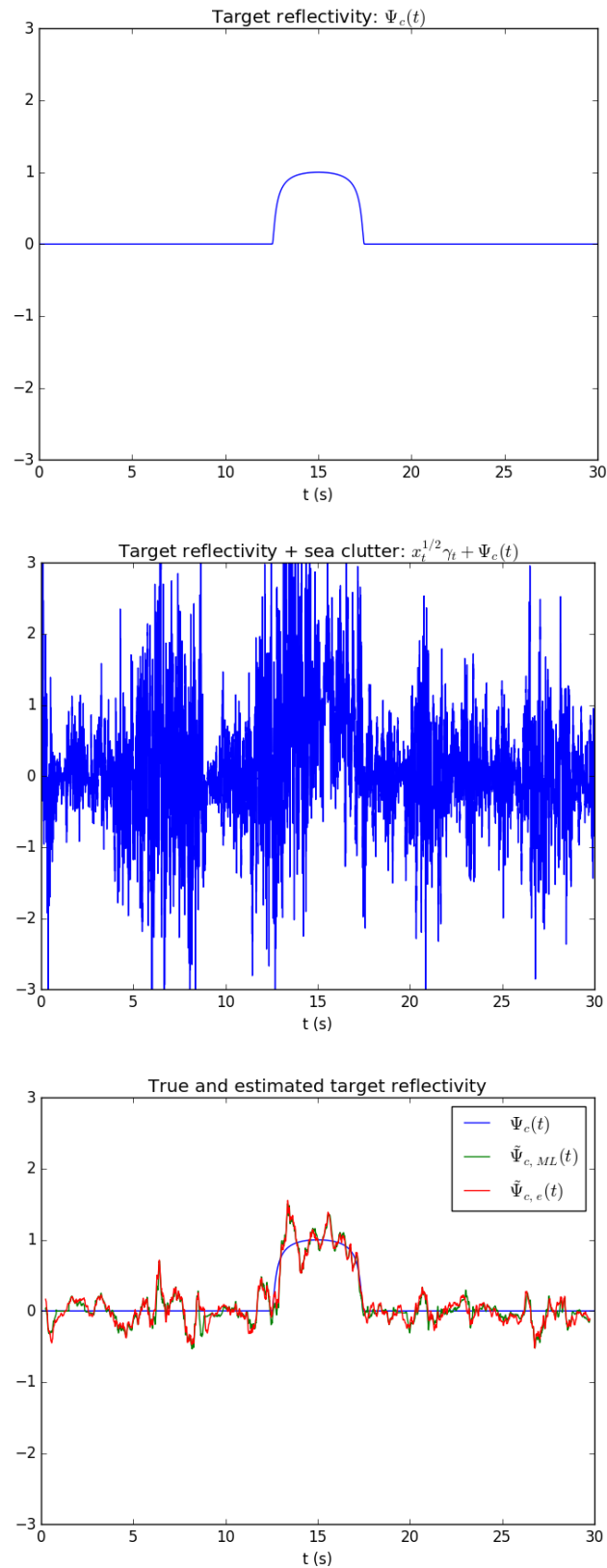
Figure 7.8: Top: target reflectivity using equation (7.59). Middle: target + sea clutter. Bottom: true and estimated target reflectivity.

Second approach: likelihood analysis

It is possible to test for the presence of an anomaly directly from the likelihood of a time series. Assume that we observe discrete time series of x_t, R_t and I_t:

$$(\tilde{x}, \tilde{R}, \tilde{I}) = \left\{ \left(\tilde{x}_k, \tilde{R}_k, \tilde{I}_k \right), k = 0, 1, \ldots, n \right\},$$

at times t_k. Under the null hypothesis H_0 (sea clutter only), we can define the likelihood function as:

$$\mathcal{L}\left(\tilde{x}, \tilde{R}, \tilde{I} \right) = p^{\infty} \prod_{k=1}^{n} p^{(k)}, \tag{7.61}$$

with

$$
\begin{aligned}
p^{\infty} &= p((x_{t_0}, R_{t_0}, I_{t_0}) = (\tilde{x}_0, \tilde{R}_0, \tilde{I}_0)) \\
&= p^{\infty}(\tilde{x}_0, \tilde{R}_0, \tilde{I}_0)
\end{aligned} \tag{7.62}
$$

and

$$p^{(k)} = p\left((x_{t_k}, R_{t_k}, I_{t_k}) = (\tilde{x}_k, \tilde{R}_k, \tilde{I}_k) \mid (x_{t_{k-1}}, R_{t_{k-1}}, I_{t_{k-1}}) = (\tilde{x}_{k-1}, \tilde{R}_{k-1}, \tilde{I}_{k-1}) \right). \tag{7.63}$$

It is possible to direcly decide that there is an anomaly if $-\mathcal{L}$ exceeds some threshold, *i.e.* if $(\tilde{x}, \tilde{R}, \tilde{I})$ is too unlikely under the null. Of course, in practice x_t must be estimated first. The non-observability of x_t cannot be solved using $\tilde{z} = \{\tilde{z}_k, k = 0, 1, \ldots, n\}$, where $\tilde{z}_k = \tilde{R}_k^2 + \tilde{I}_k^2$. Even though $\tilde{z}$ is observed, its likelihood $\mathcal{L}(\tilde{z})$ cannot simply be defined as:

$$\mathcal{L}(\tilde{z}) = p(z_{t_0} = \tilde{z}_0) \prod_{k=1}^{n} p(z_{t_k} = \tilde{z}_k \mid z_{t_{k-1}} = \tilde{z}_{k-1}). \tag{7.64}$$

Indeed, z_t is not a diffusion process, in the sense that its future does not depend only on its present state (unlike diffusions), but on the present state of two processes. This was noticed in chapter 4 when we computed its transition probabilities. The correct formulation for the likelihood of $\tilde{z}$ is:

$$\mathcal{L}(\tilde{z}) = p(z_{t_0} = \tilde{z}_0) \prod_{k=1}^{n} p(z_{t_k} = \tilde{z}_k \mid x_{t_{k-1}} = \tilde{x}_{k-1}, \gamma_{t_{k-1}} = \tilde{\gamma}_{k-1}). \tag{7.65}$$

Observing x_t, or estimating it, is again necessary. It is then possible to set a threshold on $-\mathcal{L}(\tilde{z})$ and decide that there is an anomaly if the threshold is exceeded. A drawback of this method is that x_t must be estimated and that it seems for now that the estimation of x_t in presence of a target can be done only if the target reflectivity is an additive (smoothly time-dependent) term. An advantage is that besides requiring a smooth target reflectivity, no other assumption are made on the target.

In practice, we would propose to use a sliding window Δ and compute the likelihood of $\{(x_{t_k}, R_{t_k}, I_{t_k}), \forall k \in \Delta\}$ for all positions of the window.

7.5 Conclusion

In comparison with the rest of the thesis, this closing chapter, divided into two parts, was more prospective and open to new contributions.

In the first part (sections 7.1 and 7.2), we addressed the non-observability of x_t. In section 7.1, we use the difference in timescales between x_t and γ_t to derive heuristic estimators for x_t based on squared increments of the complex reflectivity Ψ_t (equations (7.11) and (7.24)). Though the calculation is detailed for the case without a target, we assert that the estimator is also valid for slowly varying targets with additive reflectivity (see section 7.4). Synthetizing the previously derived estimators for $\mathcal{A}$, $\mathcal{B}$, α (under the hypothesis that x_t is observed), and the estimator for x_t based on the (observable) complex reflectivity Ψ_t, we proposed in section 7.2 a sequence of estimators which can be used as such on real data. Based on numerical experiments, we see that the estimation of $\mathcal{A}$ requires that we subsample the estimated x_t trajectory. We recognize that our work on the topic is incomplete and should be continued further using information theory notions to adequatly appraise the proposed sequence of estimators, and in particular its convergence as it is relooped.

In the second part, we introduced bayesian estimation of the clutter parameters (section 7.3) and target detection (section 7.4). Bayesian estimation of $\mathcal{A}$, α and $\mathcal{B}$ is straightforward because we already have computed the likelihood functions of x_t and γ_t in chapter 5. The relevance of bayesian estimation still has to be proved numerically by MCMC simulations. We presented two methods for target detection, one based on spot volatility and one based on likelihood (with two different approaches, parameter estimation and likelihood analysis). The method based on spot volatility can be easily applied. However, spot volatility estimation has slower convergence rate than integrated volatility estimation. Numerical simulations could assess whether it is viable with a realistic timestep (*e.g.* 10^{-3} s). If it is not, the method based on likelihood should be preferred.

Conclusion

Due to the complex dynamic behavior of its surface, radar remote sensing of the sea results in noise-like signals for its reflectivity, termed clutter. A first gap in the previous research on radar remote sensing of the sea surface was that Synthetic Aperture Radar (SAR) imaging does not currently take into account the dynamics of the sea in its time-integration procedure. A second gap can be stated as follows: a static statistical model for the clutter, namely the K distribution, has existed for decades. The K distribution is both derived theoretically from the random walk model, and verified empirically. However, how can a static model capture the essence of a dynamic physical phenomenon? This question can stand by itself from a purely descriptive point of view, but it is also significant from an utilitary point of view. When it comes to signal processing of sea clutter, would not it be better to have a dynamic rather than a static subtending model?

A dynamic extension to the K distribution/random walk model has been developped in the early 2000s by T. R. Field. The idea of this thesis was to approach the two prementionned gaps at the light of Field's model. Notwithstanding the significant step forward that it is, we think that Field's model as reported in [48] needed serious clarification due to a notable lack of details and to notational inconsistencies. Therefore, our first contribution has been to clear up Field's model (chapter 3), though some dark zones still remain (for me!) within the proofs. It was possible to present it concisely as a system of three stochastic differential equations depending on three parameters $\mathcal{A}$, $\mathcal{B}$ and α.

Our contribution to filling the first gap was reported in chapter 4. Solving Fokker-Planck equations, we expressed forward and backward transition probabilities for the sea clutter and proposed to use them as a tool for synchronizing data acquired from different positions at different times. We imply that this could be used in the SAR imaging time-integration. Numerical simulations were used to illustrate the spreading of transition probabilities (increasing uncertainty) as the time interval increases.

We addressed the question of estimating the parameters $\mathcal{A}$, $\mathcal{B}$ and α. At first, we proposed to use maximum likelihood (chapter 5) or volatility-based estimation (appendix A.1) for $\mathcal{A}$ and $\mathcal{B}$, and ergodicity/moment-based estimators for α (chapter 5). Afterwards, we highlighted that a key hypothesis used in the first step, namely that the radar cross section x_t is observed, was not true in practice. We derived estimators for x_t, as well as an estimator for $\mathcal{B}$, based on observable-only data in chapter 7. They enabled us to untangle the circularities and to finally propose a sequence of estimators for $\mathcal{A}, \mathcal{B}, \alpha, x_t$ and γ_t (chapter 7). Numerical simulations were systematically used to assess the performance of the estimators.

We derived stochastic differential equations for the target plus clutter complex reflectivity in a cartesian coordinate system (while Field had derived the corresponding SDE in polar coordinates). In either of the two target models under consideration, Homodyned K and Generalized K scattering, the target is parameterized by a complex constant Ψ_c. We proposed

estimators for Ψ_c based on maximum-likelihood (chapter 6) and volatility (appendix A.2). Numerical simulations were again used to assess the performance of the estimators.

These results were published in peer-reviewed journals in the form of three different papers: [122], [121] and [123] (this last one is still being peer-reviewed). They were also presented in several conferences.

Field's model was proposed in [48] for electromagnetic scattering by random media. As a particular case, we chose to use it to describe the scattering of radar waves by the sea surface. On one hand, Field's model has, by construction, greater generality than that. This means that all the results we obtained in this thesis are potentially usable in contexts other than radar remote sensing of the sea. It may be used for phase screens [70], [48], for wireless channel modelling [47], or even maybe for sound waves scattering by turbulent flows for submarine applications. On the other hand, when evaluated as a model for sea clutter, it may be noticed that Field's model lacks specificity. On one hand, the phase model for individual scatterers (Gaussian transitions), as well as the assumption that different scatterers are independent, is questionable [29]. If the phase model is questioned, so is the speckle γ_t. On the other hand, the population model which led to the radar cross section x_t is also questionable. For high-resolution radars, the assumption that there are infinitely many scatterers breaks down. Also, and possibly related to the remarks on the phase model, cyclic motions of the sea surface, like the swell, are not accounted for in Field's model. They should involve oscillations in the autocorrelation function of x_t instead of a purely decreasing exponential.

At the light of the results reported in this thesis, and of the strenghs and weaknesses of the foundation it was built on, we can now propose a number of directions for future work.

First, to rebound on the preceding remarks on Field's model, it is possible to expand in either of the two following directions: new applications and modified model. The first one consists in finding applications in which Field's model and the results of this thesis could play a role in solving unresolved problems. The second one consists in modifying Field's model to make it more adapted to the specific case of sea clutter. We think that, especially for high resolution radars, it might become necessary to remove the infinite population of scatterers hypothesis. However, this would lead to a finite random walk and to discontinuities in the complex reflectivity. The framework of stochastic differential equations driven by brownian motions would collapse and it would be necessary to introduce new tools such as jump-diffusion models (see [133] for a primer). It would also be relevant to include cyclostationarity in x_t, to model the cyclicity of sea waves. This could be done combining the BDI population model used in [48], and the sum of cosines used in [57]. It might also be compelling to take off the independent scatterers hypothesis and introduce more structure, such as a common velocity component, swell etc. Modifying Field's model would make the detailed results of this thesis obsolete, but the outline would remain. For example, if cyclicity is introduced in x_t, we could investigate if, and how, one can define transition probabilities.

Second, the first gap cited above is still largely open. Though we contributed to reducing it, there is a lot of effort to be made to include a dynamic stochastic model in SAR imaging. Focusing on the heart of SAR algorithms shall reveal their flaw and where there is room for a dynamic model (Field's model, or a more adapted one).

Third, it is possible now to apply the sequence of estimators proposed in chapter 7 to real data, and, in a way, to test the adequacy of Field's model with reality. Free access sea clutter and target reflectivity databases exist online [37], such as the IPIX Radar Database, recorded and published by McMaster University, Canada [2]. They are ideal starting points. We have also seen that the sequence of estimators relies on the estimator for x_t, which is purely heuristic and could be replaced by a better estimator. Since the non-observability of the radar cross section x_t is one of the biggest hindrance in working on real data, its study could be the central node of future reflections.

Fourth, two almost unexplored topics within the framework of Field's model are bayesianism and target detection. New research could start from the end of chapter 7, to respectively test numerically the bayesian estimation of the clutter parameters, and to further explicit the target detection scheme that we introduced.

Fifth, we have seen that estimators become better when data (and therefore information) is added: this is nothing less than *learning*. From a general perspective, machine learning techniques are quite appealing, and we think that they would find their way easily to *e.g.* anomaly detection.

Finally, we have noticed in chapter 5 that different estimators were giving nearly identical estimations from the same finite discrete time series. This fact encourages us to think of estimators as tools which extract information from data about something (*e.g.* the parameters), in which case the identical results stem from the finite amount of information in the data, and the similar capacity of the estimators to extract that information. Of course, parameter estimation is just an example, but it could be valuable to import notions from *information theory*, such as the *entropy, Kullback-Leibler risk* etc. Interestingly enough, we observe that the likelihood function and maximum likelihood parameter estimation followed us from chapter 5 until the end of the thesis, from clutter and target parameter estimation, to the likelihood function in Bayes formula and target detection: information theory is already on the doorstep (for more on the role of the likelihood function in information theory, see [28], [32], [33] and [125]). More specifically, model testing could be developed in response to the remarks about the estimation of the RCS, speckle and clutter parameters made in section 7.2.

Undoubtedly, our contribution to science has not completely enlightened the topic of electromagnetic scattering from the sea surface. *"Only that day dawns to which we are awake. There is more day to dawn."*

Appendix A

Volatility-based estimation

In chapter 5, we presented the estimation of the clutter parameters of Field's model, $\mathcal{A}$, $\mathcal{B}$ and α. In chapter 6, we presented the estimation of the target parameters $\Psi_c^{(R)}$, $\Psi_c^{(I)}$. Under the assumption that α is known, we estimated $\mathcal{A}$ and $\mathcal{B}$ by maximum likelihood (ML). Under the assumption that $\mathcal{A}$, $\mathcal{B}$ and α are known, we also estimated $\Psi_c^{(R)}$ and $\Psi_c^{(I)}$ by ML. Euler-Maruyama's approximation for the transition probabilities lead to explicit expressions for the estimators in terms of the observed data. However, the derivations of the estimators is relatively long by this approach.

Thanks to a collaborative work with Randolf Altmeyer (Humboldt University of Berlin), we can propose a simpler approach for estimating $\mathcal{A}$, $\mathcal{B}$, $\Psi_c^{(R)}$ and $\Psi_c^{(I)}$ based on volatility estimation.

A.1 Volatility-based estimation of $\mathcal{A}$ and $\mathcal{B}$

A.1.1 Estimation of $\mathcal{A}$

We know from chapter 3 that:

$$\mathrm{d}x_t = \mathcal{A}(1 - x_t)\mathrm{d}t + \left(2\frac{\mathcal{A}}{\alpha}x_t\right)^{\frac{1}{2}}\mathrm{d}W_t^{(x)}. \tag{A.1}$$

The differential of the quadratic variation of x_t is:

$$\mathrm{d}\langle x \rangle_t = 2\frac{\mathcal{A}}{\alpha}x_t\mathrm{d}t, \tag{A.2}$$

which implies

$$\langle x \rangle_t = \int_0^t 2\frac{\mathcal{A}}{\alpha}x_r\mathrm{d}r. \tag{A.3}$$

Since one possible estimator of $\langle x \rangle_t$ is:

$$\sum_{k=1}^n (x_{t_k} - x_{t_{k-1}})^2$$

and one possible estimator of $\int_0^t 2\frac{\mathcal{A}}{\alpha}x_r\mathrm{d}r$ is:

$$\sum_{k=1}^n 2\frac{\mathcal{A}}{\alpha}x_{t_k}\Delta t,$$

we can define the following estimator for $\mathcal{A}$:

$$\tilde{\mathcal{A}}_{vol} = \frac{\alpha \sum_{k=1}^{n}(x_{t_k} - x_{t_{k-1}})^2}{2\Delta t \sum_{k=1}^{n} x_{t_k}}. \tag{A.4}$$

A.1.2 Estimation of $\mathcal{B}$

We know from chapter 3 that:

$$\mathrm{d}\gamma_t^{(R)} = -\frac{1}{2}\mathcal{B}\gamma_t^{(R)}\mathrm{d}t + \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\mathrm{d}W_t^{(R)}. \tag{A.5}$$

The differential of the quadratic variation of $\gamma_t^{(R)}$ is:

$$\mathrm{d}\langle\gamma^{(R)}\rangle_t = \frac{\mathcal{B}}{2}\mathrm{d}t, \tag{A.6}$$

which implies

$$\langle\gamma^{(R)}\rangle_t = \int_0^t \frac{\mathcal{B}}{2}\mathrm{d}r = \mathcal{B}\frac{t}{2} = \mathcal{B}\frac{n\Delta t}{2}. \tag{A.7}$$

Since $\sum_{k=1}^{n}(\gamma_{t_k}^{(R)} - \gamma_{t_{k-1}}^{(R)})^2$ is an estimator of $\langle\gamma^{(R)}\rangle_t$, we have the following estimator for $\mathcal{B}$:

$$\tilde{\mathcal{B}}_{vol} = \frac{2}{n\Delta t}\sum_{k=1}^{n}(\gamma_{t_k}^{(R)} - \gamma_{t_{k-1}}^{(R)})^2. \tag{A.8}$$

The same procedure can be applied independently to $\gamma_t^{(I)}$, with the $\gamma_{t_k}^{(I)}$ replacing the $\gamma_{t_k}^{(R)}$.

A.1.3 Numerical experiments

To assess the performance of the volatility-based estimators for $\mathcal{A}$ and $\mathcal{B}$ in realistic configurations, we conduct numerical experiments. We set values for the parameters $\mathcal{A}$ and $\mathcal{B}$, simulate many trajectories of x_t, $\gamma_t^{(R)}$, and $\gamma_t^{(I)}$, and estimate $\mathcal{A}$ and $\mathcal{B}$ for each trajectory. We set $\alpha = 1$ in all the simulations.

For each value of $\mathcal{A}$ in the interval $[0.1, 10]$ Hz (with a step of 0.1 Hz), we generate $N = 1000$ trajectories of x_t, $\{\tilde{x}^{(i)}, i = 0, 1, \ldots, N\}$, using Milstein's scheme (see [62]). For all i, $\tilde{x}^{(i)} = \{\tilde{x}_k^{(i)}, k = 1, 2 \ldots, n\}$. The simulation timestep is $\hat{\Delta}t = 10^{-4}$ s. The observations are at times t_k with constant timestep $\Delta t = t_k - t_{k-1}$, either 10^{-4} or 10^{-3} s. This is to appreciate the effect of changing the timestep. $\Delta t = 10^{-3}$ s is the right order of magnitude for many radars, but $\Delta t = 10^{-4}$ s is also achievable. For each trajectory $\tilde{x}^{(i)}$, $\mathcal{A}$ is estimated with formula (A.4). The estimation bias $b(\mathcal{A})$ and standard deviation $\sigma(\mathcal{A})$ are calculated from the estimations $\tilde{\mathcal{A}}_1, \tilde{\mathcal{A}}_2, \ldots, \tilde{\mathcal{A}}_N$. For comparison, the same is done for ML estimation, where the transition probabilities are approximated by Gaussian random variables according to Euler-Maruyama scheme (as in chapter 5).

The same approach is carried out with $\mathcal{B}$. For each value of $\mathcal{B}$ in the interval $[10, 1000]$ Hz with a step of 10 Hz, we simulate $N = 1000$ trajectories $\{\tilde{\gamma}^{(i)}, i = 0, 1, \ldots, N\}$ of γ_t using Euler-Maruyama's scheme. For all i, $\mathcal{B}$ is estimated from the real and imaginary parts of $\tilde{\gamma}^{(i)}$ using equation (A.8) and the average estimation is retained. Again, we compare the results with the ML estimator with Euler-Maruyama's approximation for the transition probabilities as in chapter 5.

Based on the numerical experiments, we observed that for $\mathcal{A}$, the two estimators (volatility-based and ML) have about the same standard deviation (both for $\Delta t = 10^{-3}$ s and $\Delta t = 10^{-4}$ s). They are significantly biased but in opposite directions for $\Delta t = 10^{-3}$ s, and the bias is almost zero for $\Delta t = 10^{-4}$ s. For $\mathcal{B}$, the volatility-based estimator is slightly less biased, but has a larger standard deviation.

As explained in chapter 5, a relevant way to compare the two estimators is to compute their root mean square error (RMSE) after debiasing. We remind the debiasing procedure: let for example $\mathcal{A}_b(\mathcal{A})$ be a biased estimator of $\mathcal{A}$ with bias $b(\mathcal{A})$. To debias the estimator, we solve the following equation in $\mathcal{A}_{db}(\mathcal{A})$:

$$\mathcal{A}_b(\mathcal{A}) = \mathcal{A}_{db}(\mathcal{A}) + b(\mathcal{A}_{db}(\mathcal{A})), \tag{A.9}$$

and obtain the debiased estimator $\mathcal{A}_{db}(\mathcal{A})$. For fixed $\mathcal{A}$, the RMSE is then computed from the N trajectories as:

$$rmse(\mathcal{A}_{db})^2 = \frac{1}{N} \sum_{i=1}^{N} (\mathcal{A}_{db,\,i} - \mathcal{A})^2. \tag{A.10}$$

Applying the bias correction and computing the RMSE, we obtain the results in figure A.1 for $\mathcal{A}$ and $\mathcal{B}$. It is remarkable that the volatility-based and ML estimators have almost identical RMSE. The larger bias of ML transforms into additional standard deviation when the bias correction is applied, such that overall the two estimators have identical performance.

A.2   Volatility-based estimation of target parameters

In chapter 6 section 6.1, we derived the SDEs for the multidimensional process $Y_t :=$ $\begin{bmatrix} x_t & R_t & I_t \end{bmatrix}^\top$, where $R_t = \mathrm{Re}(\Psi_t)$ and $I_t = \mathrm{Im}(\Psi_t)$, and where:

$$\Psi_t = \Psi_t^{(HK)} = \Psi_c^{(R)} + i\Psi_c^{(I)} + x_t^{1/2}\gamma_t \tag{A.11}$$

for HK scattering, and

$$\Psi_t = \Psi_t^{(GK)} = \left(\Psi_c^{(R)} + i\Psi_c^{(I)}\right)\eta x_t + x_t^{1/2}\gamma_t \tag{A.12}$$

for GK scattering. The SDEs are respectively equations (6.12) and (6.17) for HK and GK scattering. In section 6.2, we proposed ML estimators for $\Psi_c^{(R)}$ and $\Psi_c^{(I)}$ using Euler-Maruyama's scheme to approximate the transition probabilities of $\begin{bmatrix} x_t & R_t & I_t \end{bmatrix}^\top$ by Gaussian distributions. After long calculations mainly developed in appendices D.1.2 and D.2.2, we obtained equations (6.36) and (6.50) for the estimators in HK and GK scattering respectively. In this section, we derive volatility-based estimators for $\Psi_c^{(R)}$ and $\Psi_c^{(I)}$ for HK and GK scattering. We carry out numerical simulations to assess these new estimators and compare them to the ML estimators of chapter 6.

A.2.1   Estimating Ψ_c in HK scattering

In the case of HK scattering, we have seen that the squared-volatility of the process Y_t is (equation (6.29)):

$$\sigma_t = \begin{bmatrix} \frac{2\mathcal{A}x_t}{\alpha} & \frac{\mathcal{A}(R_t - \Psi_c^{(R)})}{\alpha} & \frac{\mathcal{A}(I_t - \Psi_c^{(I)})}{\alpha} \\ \frac{\mathcal{A}(R_t - \Psi_c^{(R)})}{\alpha} & \frac{\mathcal{A}}{2\alpha}\frac{(R_t - \Psi_c^{(R)})^2}{x_t} + \frac{\mathcal{B}x_t}{2} & \frac{\mathcal{A}}{2\alpha}\frac{(R_t - \Psi_c^{(R)})(I_t - \Psi_c^{(I)})}{x_t} \\ \frac{\mathcal{A}(I_t - \Psi_c^{(I)})}{\alpha} & \frac{\mathcal{A}}{2\alpha}\frac{(R_t - \Psi_c^{(R)})(I_t - \Psi_c^{(I)})}{x_t} & \frac{\mathcal{A}}{2\alpha}\frac{(I_t - \Psi_c^{(I)})^2}{x_t} + \frac{\mathcal{B}x_t}{2} \end{bmatrix}. \tag{A.13}$$
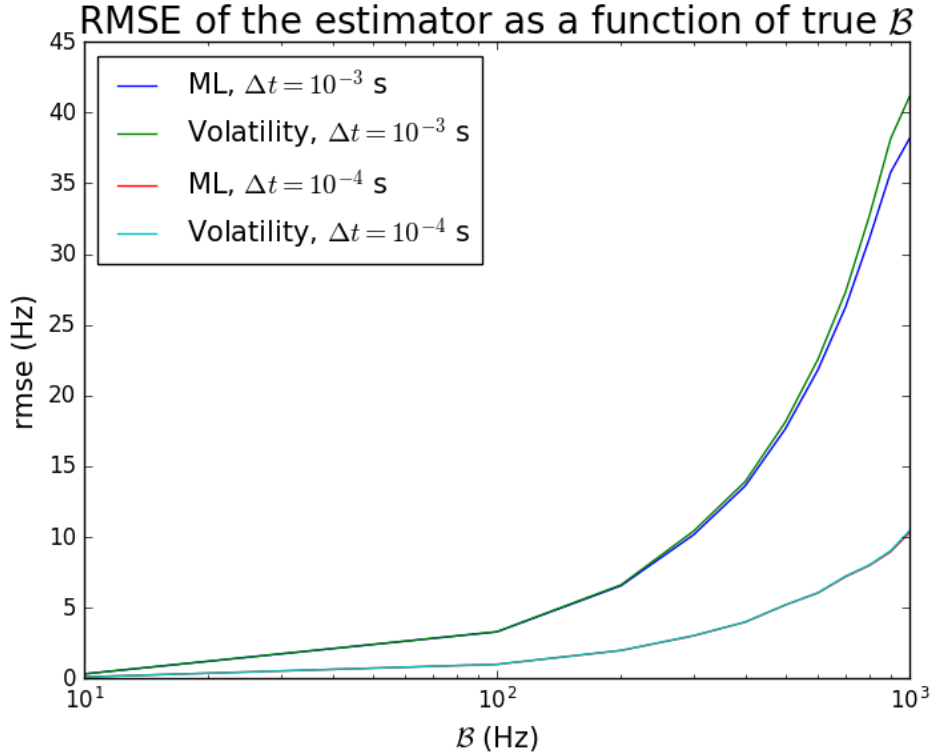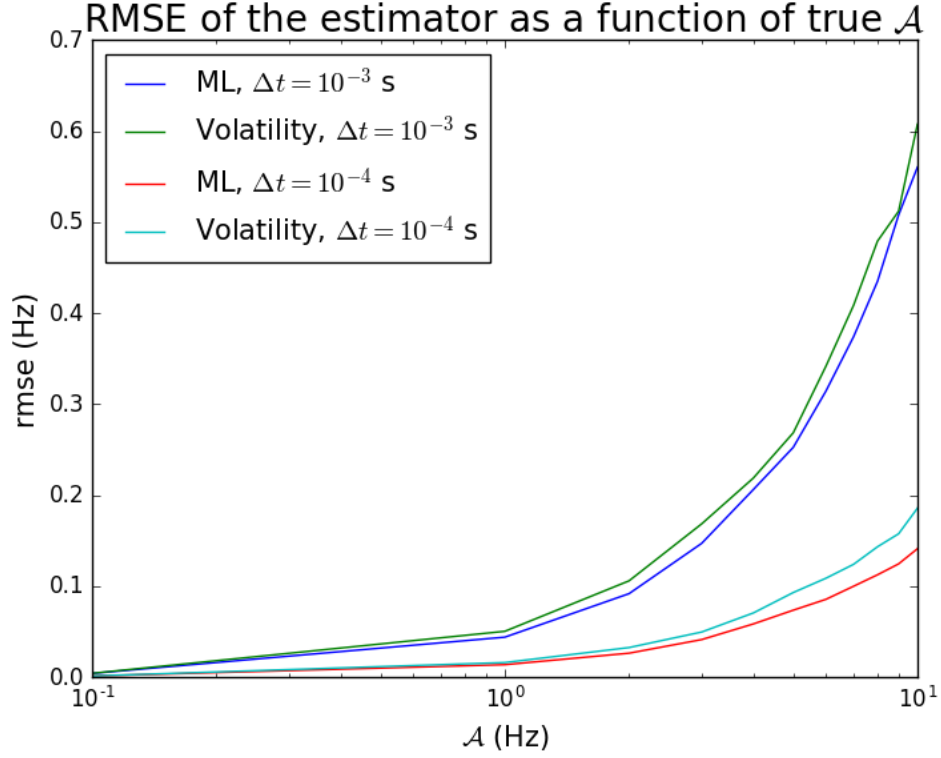
Figure A.1: RMSE of the estimators for $\mathcal{A}$ (up) and for $\mathcal{B}$ (down) as a function of true $\mathcal{A}$ and $\mathcal{B}$. Two estimators are compared: the volatility-based estimator and the ML estimator with Euler's approximation for the transition probabilities.

The quadratic variation of Y_t is $\langle Y \rangle_t = \int_0^t \sigma_r dr = IV_t$, where IV is the integrated volatility. This means that for $t > 0$, we have

$$\int_0^t \sigma_r^{21} \mathrm{d}r = \frac{\mathcal{A}}{\alpha} \int_0^t \left[R_r - \Psi_c^{(R)} \right] \mathrm{d}r = \frac{\mathcal{A}}{\alpha} \int_0^t R_r \mathrm{d}r - \frac{\mathcal{A}t}{\alpha} \Psi_c^{(R)}, \qquad (A.14)$$

from which we get:

$$\Psi_c^{(R)} = \frac{1}{t}\int_0^t R_r \mathrm{d}r - \frac{\alpha}{t\mathcal{A}}\int_0^t \sigma_r^{21} dr. \tag{A.15}$$

Similarly,

$$\Psi_c^{(I)} = \frac{1}{t}\int_0^t I_r \mathrm{d}r - \frac{\alpha}{t\mathcal{A}}\int_0^t \sigma_r^{31} dr. \tag{A.16}$$

Assume now that we know $\mathcal{A}$ and α and that we observe Y_{t_k} (and thus R_{t_k}, I_{t_k}, x_{t_k}) for $0 = t_0 < t_1 < \cdots < t_n = t$, $t_k = k\Delta t$, $\Delta t = t/n$. This allows us to form the estimator

$$RV_t^n = \sum_{k=1}^n \left(Y_{t_k} - Y_{t_{k-1}}\right)\left(Y_{t_k} - Y_{t_{k-1}}\right)^\top. \tag{A.17}$$

RV is the realized volatility and it is an estimator of the integrated volatility $IV = \int_0^t \sigma_r dr$. In particular,

$$RV_{n,t}^{21} = \sum_{k=1}^n \left(x_{t_k} - x_{t_{k-1}}\right)\left(R_{t_k} - R_{t_{k-1}}\right) \tag{A.18}$$

is an estimator of $\int_0^t \sigma_r^{21}\mathrm{d}r$ and

$$RV_{n,t}^{31} = \sum_{k=1}^n \left(x_{t_k} - x_{t_{k-1}}\right)\left(I_{t_k} - I_{t_{k-1}}\right) \tag{A.19}$$

is an estimator of $\int_0^t \sigma_r^{31}\mathrm{d}r$. Since $\sum_{k=1}^n R_{t_k}\Delta t$ is an estimator of $\int_0^t R_r \mathrm{d}r$ and $\sum_{k=1}^n I_{t_k}\Delta t$ an estimator of $\int_0^t I_r \mathrm{d}r$, by equations (A.15) and (A.16) we get the following estimators for $\Psi_c^{(R)}$ and $\Psi_c^{(I)}$:

$$\tilde{\Psi}_{c,\,vol}^{(R)} = \frac{1}{t}\sum_{k=1}^n R_{t_k}\Delta t - \frac{\alpha}{t\mathcal{A}}\sum_{k=1}^n \left(x_{t_k} - x_{t_{k-1}}\right)\left(R_{t_k} - R_{t_{k-1}}\right), \tag{A.20}$$

and

$$\hat{\Psi}_{c,\,vol}^{(I)} = \frac{1}{t}\sum_{k=1}^n I_{t_k}\Delta t - \frac{\alpha}{t\mathcal{A}}\sum_{k=1}^n \left(x_{t_k} - x_{t_{k-1}}\right)\left(I_{t_k} - I_{t_{k-1}}\right). \tag{A.21}$$

A.2.2 Estimating Ψ_c in GK scattering

For GK scattering, the procedure is about the same. We have seen that in that case the squared-volatility of the process Y_t is (equation (6.42)):

$$\sigma_t = \begin{bmatrix} \frac{2\mathcal{A}x_t}{\alpha} & \frac{\mathcal{A}(R_t + \Psi_c^{(R)}\eta x_t)}{\alpha} & \frac{\mathcal{A}(I_t + \Psi_c^{(I)}\eta x_t)}{\alpha} \\ \frac{\mathcal{A}(R_t + \Psi_c^{(R)}\eta x_t)}{\alpha} & \frac{\mathcal{A}}{2\alpha}\frac{(R_t + \Psi_c^{(R)}\eta x_t)^2}{x_t} + \frac{\mathcal{B}x_t}{2} & \frac{\mathcal{A}}{2\alpha}\frac{(R_t + \Psi_c^{(R)}\eta x_t)(I_t + \Psi_c^{(I)}\eta x_t)}{x_t} \\ \frac{\mathcal{A}(I_t + \Psi_c^{(I)}\eta x_t)}{\alpha} & \frac{\mathcal{A}}{2\alpha}\frac{(R_t + \Psi_c^{(R)}\eta x_t)(I_t + \Psi_c^{(I)}\eta x_t)}{x_t} & \frac{\mathcal{A}}{2\alpha}\frac{(I_t + \Psi_c^{(I)}\eta x_t)^2}{x_t} + \frac{\mathcal{B}x_t}{2} \end{bmatrix}. \tag{A.22}$$

For $t > 0$, we have

$$\begin{aligned}
\int_0^t \sigma_r^{21}\mathrm{d}r &= \int_0^t \frac{\mathcal{A}(R_r + \Psi_c^{(R)}\eta x_r)}{\alpha}\mathrm{d}r = \frac{\mathcal{A}}{\alpha}\int_0^t R_r \mathrm{d}r + \frac{\mathcal{A}}{\alpha}\int_0^t \Psi_c^{(R)}\eta x_r \mathrm{d}r \\
&= \frac{\mathcal{A}}{\alpha}\int_0^t R_r \mathrm{d}r + \Psi_c^{(R)}\frac{\mathcal{A}\eta}{\alpha}\int_0^t x_r \mathrm{d}r
\end{aligned} \tag{A.23}$$

193

from which we get:

$$\Psi_c^{(R)} = \frac{\frac{\alpha}{\eta\mathcal{A}} \int_0^t \sigma_r^{21} dr - \frac{1}{\eta} \int_0^t R_r \mathrm{d}r}{\int_0^t x_r \mathrm{d}r}. \tag{A.24}$$

Similarly,

$$\Psi_c^{(I)} = \frac{\frac{\alpha}{\eta\mathcal{A}} \int_0^t \sigma_r^{31} dr - \frac{1}{\eta} \int_0^t I_r \mathrm{d}r}{\int_0^t x_r \mathrm{d}r}. \tag{A.25}$$

Under the same assumptions as for HK scattering, we form the estimator

$$RV_t^n = \sum_{k=1}^n \left(Y_{t_k} - Y_{t_{k-1}}\right) \left(Y_{t_k} - Y_{t_{k-1}}\right)^\top. \tag{A.26}$$

for the integrated volatility $IV = \int_0^t \sigma_r dr$. By equations (A.24) and (A.25), and using the same estimator as for HK scattering for the various integrals, we get the following estimators for $\Psi_c^{(R)}$ and $\Psi_c^{(I)}$:

$$\tilde{\Psi}_{c,\,vol}^{(R)} = \frac{\frac{\alpha}{\eta\mathcal{A}} \sum_{k=1}^n \left(x_{t_k} - x_{t_{k-1}}\right)\left(R_{t_k} - R_{t_{k-1}}\right) - \frac{1}{\eta} \sum_{k=1}^n R_{t_k} \Delta t}{\sum_{k=1}^n x_{t_k} \Delta t}, \tag{A.27}$$

and

$$\tilde{\Psi}_{c,\,vol}^{(I)} = \frac{\frac{\alpha}{\eta\mathcal{A}} \sum_{k=1}^n \left(x_{t_k} - x_{t_{k-1}}\right)\left(I_{t_k} - I_{t_{k-1}}\right) - \frac{1}{\eta} \sum_{k=1}^n I_{t_k} \Delta t}{\sum_{k=1}^n x_{t_k} \Delta t}. \tag{A.28}$$

A.2.3 Numerical experiments

We do numerical experiments to assess the volatility-based estimators that we derived in sections A.2.1 and A.2.2. We also want to compare them to the ML and the ergodicity-based estimators of chapter 6 (see equation (6.54) for the ergodicity-based estimator). As in section 6.3.2, to assess the estimators we simulate many trajectories of Ψ_t with know target constant Ψ_c, estimate Ψ_c for each trajectory, and compute the estimation bias and standard deviation. For all simulations, we set $\mathcal{A} = 1$ Hz, $\mathcal{B} = 100$ Hz, $\alpha = 1$, $\eta = 1$ and $\Delta t = 10^{-3}$ s. Also, $\Psi_c^{(I)}$ is set to zero such that all the power of the target goes to the real part. For both HK and GK scattering, we explore the dependence of the estimation bias and standard deviation to the duration of the trajectories (from 0.1 to 10 s) and to the intensity of the target (its squared-modulus), from 0.1 to 100. The intensity is set to 10 when the duration is explored and the duration is set to 1 s when the intensity is explored. 1000 trajectories are generated for each scenario, using as usual Euler-Maruyama scheme for R_t and I_t, and Milstein's scheme for x_t.

Figure A.2 represents the results for HK scattering. We observe that overall, the volatility-based estimator (equation (A.20)) is able to retrieve the right parameters with almost no bias. However, it has estimation standard deviations significantly larger than the ML and ergodicity-based estimators. This is especially true for small durations and target intensities. Visually, the variance becomes reasonably small for durations larger than 1 s and intensity larger than 10. It is somehow intuitive that the volatility-based estimator should not be relevant for estimating Ψ_c in the HK scattering case, which would explain that it has low performance. Indeed, Ψ_c is then just a shift of the reflectivity and should not affect the volatility.

For GK scattering, the results are represented in figure A.3. We omitted the ergodicity-based estimator voluntarily because it would hinder the comparison of the volatility-based
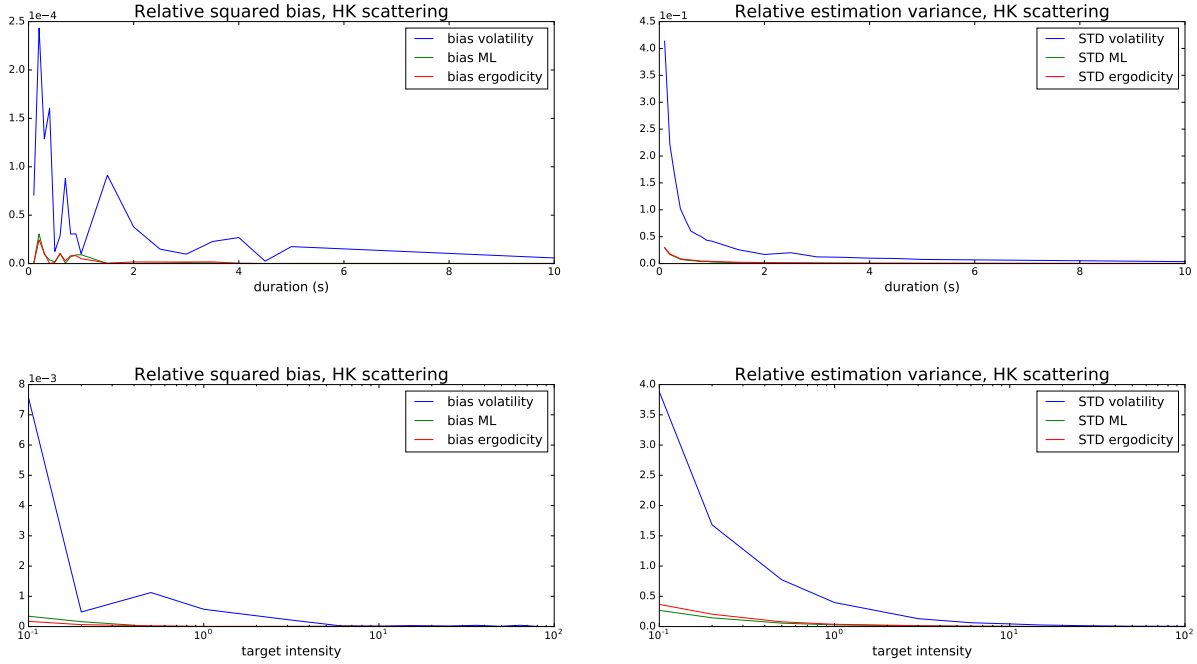
Figure A.2: Relative estimation squared bias and variance of the target in HK scattering for the volatility-based estimator (equation (A.20)). 1000 trajectories are computed, with $\mathcal{A} = 1$ Hz, $\mathcal{B} = 100$ Hz and $\alpha = 1$. Up: dependence to trajectory duration with $\Psi_c = \sqrt{10}$. Down: dependence to target intensity with a duration of 1 s.

and ML estimators. Based on figure 6.3 of chapter 6 and figure A.3 here, the ergodicity-based estimator has negligible bias but a standard deviation much higher than the ML and volatility-based estimators. The volatility-based estimator has negligible bias. Its standard deviation is greater than the ML estimator, but as we said, much lower than the ergodicity-based estimator. In terms of orders of magnitude, the ML and volatility-based estimators have similar performances in GK scattering. Contrary to HK scattering, here Ψ_c enters the volatility because it is multiplied by x_t, which explains the results. An advantage of the volatility-based estimator, besides its simplicity of use, is that it does not generate outliers like the ML estimator (see chapter 6). Those outliers, even though most of them have been filtered out, explain why the standard deviation of the ML estimator looks spiky in figure A.3.
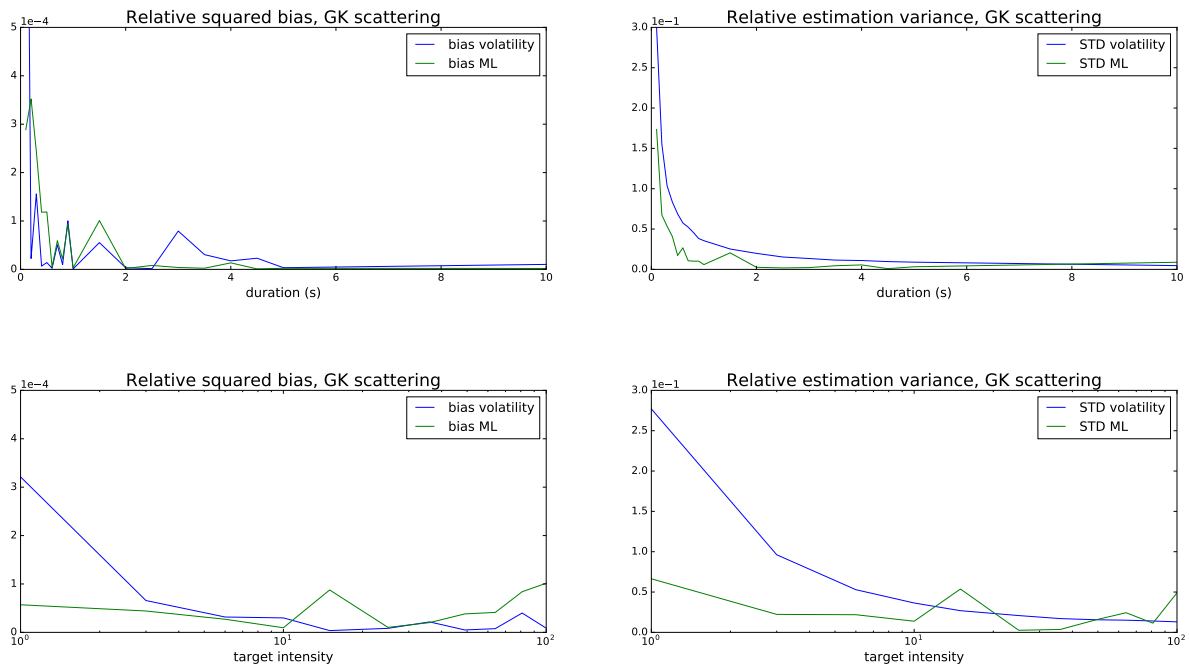
Figure A.3: Relative estimation squared bias and variance of the target in GK scattering for the volatility-based estimator (equation (A.27)). 1000 trajectories are computed, with $\mathcal{A} = 1$ Hz, $\mathcal{B} = 100$ Hz and $\alpha = 1$. Up: dependence to trajectory duration with $\Psi_c = \sqrt{10}$. Down: dependence to target intensity with a duration of 1 s.

Appendix B

Complements to chapter 4

The aim of this appendix is to show that we can compute:

$$p\left(x_t^{1/2}\gamma_t^{(R)} = x|x_0 = y,\, \gamma_0^{(R)} = z\right) \tag{B.1}$$

as the distribution of the product of 2 independent random variables:

$$p\left(x_t^{1/2}\gamma_t^{(R)} = x|x_0 = y,\, \gamma_0^{(R)} = z\right) = \int_0^{+\infty} p\left(\gamma_t^{(R)} = x/u \mid x_0 = y,\, \gamma_0^{(R)} = z\right)$$
$$p\left(x_t^{1/2} = u \mid x_0 = y,\, \gamma_0^{(R)} = z\right)\frac{1}{u}du,$$

that is to show that the product and the conditioning commute.

We know that $p\left(x_t^{1/2} = .|x_0 = x\right) = p\left(x_t^{1/2} = .|x_0 = x,\, \gamma_0^{(R)} = y\right)$ and that:

$$p\left(\gamma_t^{(R)} = .|\gamma_0^{(R)} = y\right) = p\left(\gamma_t^{(R)} = .|x_0 = x,\, \gamma_0^{(R)} = y\right). \tag{B.2}$$

We would like to compute $p\left(x_t^{1/2}\gamma_t^{(R)} = .|x_0 = x,\, \gamma_0^{(R)} = y\right)$. We show in section 4.3 that:

$$p\left(x_t^{1/2} = z,\, \gamma_t^{(R)} = w|x_0 = x,\, \gamma_0^{(R)} = y\right) = p\left(x_t^{1/2} = z|x_0 = x\right)p\left(\gamma_t^{(R)} = w|\gamma_0^{(R)} = y\right). \tag{B.3}$$

Let G be the C^1-diffeomorphism:

$$\begin{aligned} G: \quad &\mathbb{R}^2 \quad \to \mathbb{R}^2 \\ &(x,y) \quad \mapsto (xy, x) \end{aligned}$$

Let $cond\left(x_t^{1/2}, \gamma_t^{(R)}\right) = \left(cond\left(x_t^{1/2}\right), cond\left(\gamma_t^{(R)}\right)\right)$ be a random vector with the distribution:

$$p\left(cond\left(x_t^{1/2}, \gamma_t^{(R)}\right) = (z, w)\right) = p\left(x_t^{1/2} = z,\, \gamma_t^{(R)} = w|x_0 = x,\, \gamma_0^{(R)} = y\right). \tag{B.4}$$

We can show easily by integration that $cond\left(x_t^{1/2}\right)$ is a random variable with distribution $p\left(cond\left(x_t^{1/2}\right) = z\right) = p\left(x_t^{1/2} = z|x_0 = x\right)$ and that $cond\left(\gamma_t^{(R)}\right)$ is a random variable

with distribution $p\left(cond\left(\gamma_t^{(R)}\right) = w\right) = p\left(\gamma_t^{(R)} = w|\gamma_0^{(R)} = y\right)$. Moreover, $cond\left(x_t^{1/2}\right)$ and $cond\left(\gamma_t^{(R)}\right)$ are independent.

From the commutativity relation (4.5), we get:

$$p\left(\left(x_t^{1/2}\gamma_t^{(R)}, x_t^{1/2}\right) = (u, v)|x_0 = x, \gamma_0^{(R)} = y\right) = p\left(\left(cond\left(x_t^{1/2}\right) cond\left(\gamma_t^{(R)}\right)\right) = (u, v)\right),$$

and by integration:

$$p\left(x_t^{1/2}\gamma_t^{(R)} = u|x_0 = x, \gamma_0^{(R)} = y\right) = p\left(cond\left(x_t^{1/2}\right) cond\left(\gamma_t^{(R)}\right) = u\right). \qquad \text{(B.5)}$$

This last equation together with the independence of $cond\left(x_t^{1/2}\right)$ and $cond\left(\gamma_t^{(R)}\right)$ and the knowledge of their distribution justifies the computations to obtain formula (4.42).

Appendix C

Complements to chapter 5

C.1   Proof of equation (5.33)

For the estimation of $\mathcal{A}$, we have:

$$\mathcal{L}(\mathcal{A}, \alpha) = p(x_{t_0} = \tilde{x}_0) \prod_{i=1}^{n} p(x_{t_i} = \tilde{x}_i \mid x_{t_{i-1}} = \tilde{x}_{i-1}), \tag{C.1}$$

with

$$p(x_{t_0} = \tilde{x}_0) = \frac{\alpha^\alpha \tilde{x}_0^{\alpha-1} e^{-\alpha \tilde{x}_0}}{\Gamma(\alpha)}, \tag{C.2}$$

and with Euler's approximation for the transition probabilities:

$$p(x_{t_i} = \tilde{x}_i \mid x_{t_{i-1}} = \tilde{x}_{i-1}) = \frac{\sqrt{\alpha}}{\sqrt{4\pi \tilde{x}_{i-1} \mathcal{A} \Delta t}} e^{-\frac{\alpha(\tilde{x}_i - \mathcal{A}\Delta t - (1 - \mathcal{A}\Delta t)\tilde{x}_{i-1})^2}{4\mathcal{A}\Delta t \tilde{x}_{i-1}}}. \tag{C.3}$$

It is assumed here that α is known. For simplicity of notation, we write $\mathcal{L}(\mathcal{A}, \alpha) = \mathcal{L}(\mathcal{A})$. Taking the log of the likelihood function, we have:

$$
\begin{aligned}
\ln \mathcal{L}(\mathcal{A}) &= \ln\left(\frac{\alpha^\alpha \tilde{x}_0^{\alpha-1} e^{-\alpha \tilde{x}_0}}{\Gamma(\alpha)}\right) + \sum_{i=1}^{n} \ln(\sqrt{\alpha}) - \sum_{i=1}^{n} \ln\left(\sqrt{4\pi \tilde{x}_{i-1} \mathcal{A} \Delta t}\right) \\
&\quad - \sum_{i=1}^{n} \frac{\alpha\left(\tilde{x}_i - \mathcal{A}\Delta t - (1 - \mathcal{A}\Delta t)\tilde{x}_{i-1}\right)^2}{4\mathcal{A}\Delta t \tilde{x}_{i-1}} \\
\Leftrightarrow \ln \mathcal{L}(\mathcal{A}) &= \ln\left(\frac{\alpha^\alpha \tilde{x}_0^{\alpha-1} e^{-\alpha \tilde{x}_0}}{\Gamma(\alpha)}\right) + \frac{n \ln \alpha}{2} - \sum_{i=1}^{n} \ln\left(\sqrt{4\pi \tilde{x}_{i-1} \Delta t}\right) - \frac{n \ln \mathcal{A}}{2} \quad \text{(C.4)} \\
&\quad - \sum_{i=1}^{n} \frac{\alpha\left(\tilde{x}_i - \mathcal{A}\Delta t - (1 - \mathcal{A}\Delta t)\tilde{x}_{i-1}\right)^2}{4\mathcal{A}\Delta t \tilde{x}_{i-1}}.
\end{aligned}
$$

Taking the derivative with respect to $\mathcal{A}$, the first 3 terms disappear and we get:

$$\frac{\partial}{\partial \mathcal{A}} \ln \mathcal{L}(\mathcal{A}) = -\frac{\partial}{\partial \mathcal{A}} \frac{n \ln \mathcal{A}}{2} - \frac{\partial}{\partial \mathcal{A}} \sum_{i=1}^{n} \frac{\alpha \left(\tilde{x}_i - \mathcal{A}\Delta t - (1 - \mathcal{A}\Delta t)\tilde{x}_{i-1}\right)^2}{4\mathcal{A}\Delta t \tilde{x}_{i-1}}$$

$$= -\frac{n}{2}\frac{1}{\mathcal{A}} - \frac{\partial}{\partial \mathcal{A}} \sum_{i=1}^{n} \frac{\alpha}{4\mathcal{A}\Delta t \tilde{x}_{i-1}} \left[(\tilde{x}_i - \tilde{x}_{i-1}) + \mathcal{A}\Delta t(\tilde{x}_{i-1} - 1)\right]^2 \qquad (C.5)$$

$$= -\frac{n}{2}\frac{1}{\mathcal{A}} - \frac{\partial}{\partial \mathcal{A}} \sum_{i=1}^{n} \frac{\alpha}{4\mathcal{A}\Delta t \tilde{x}_{i-1}} [(\tilde{x}_i - \tilde{x}_{i-1})^2 + \mathcal{A}^2 \Delta t^2 (\tilde{x}_{i-1} - 1)^2$$

$$+\ 2\mathcal{A}\Delta t (\tilde{x}_i - \tilde{x}_{i-1})(\tilde{x}_{i-1} - 1)]. \qquad (C.6)$$

After differentiation, the third term in the sum disappears (it does not depend on $\mathcal{A}$) and we get:

$$\frac{\partial}{\partial \mathcal{A}} \ln \mathcal{L}(\mathcal{A}) = -\frac{n}{2}\frac{1}{\mathcal{A}} + \sum_{i=1}^{n} \frac{\alpha(\tilde{x}_i - \tilde{x}_{i-1})^2}{4\tilde{x}_{i-1}\Delta t}\frac{1}{\mathcal{A}^2} - \sum_{i=1}^{n} \frac{\alpha(\tilde{x}_{i-1} - 1)^2 \Delta t}{4\tilde{x}_{i-1}}. \qquad (C.7)$$

The optimization condition is:

$$\frac{\partial}{\partial \mathcal{A}} \ln \mathcal{L}(\mathcal{A}) = 0,$$

$$\mathcal{A}^2 \frac{\partial}{\partial \mathcal{A}} \ln \mathcal{L}(\mathcal{A}) = 0, \qquad (C.8)$$

which states that:

$$-\sum_{i=1}^{n} \frac{\alpha(\tilde{x}_{i-1} - 1)^2 \Delta t}{4\tilde{x}_{i-1}} \mathcal{A}^2 - \frac{n}{2}\mathcal{A} + \sum_{i=1}^{n} \frac{\alpha(\tilde{x}_i - \tilde{x}_{i-1})^2}{4\tilde{x}_{i-1}\Delta t} = 0. \qquad (C.9)$$

We have proven the first line of equation (5.33).

Let us now prove the second line of equation (5.33). For $\mathcal{B}$, the likelihood function is:

$$\mathcal{L}(\mathcal{B}) = p\left(\gamma_{t_0}^{(R)} = \tilde{\gamma}_0^{(R)}\right) \prod_{i=1}^{n} p\left(\gamma_{t_i}^{(R)} = \tilde{\gamma}_i^{(R)} \mid \gamma_{t_{i-1}}^{(R)} = \tilde{\gamma}_{i-1}^{(R)}\right), \qquad (C.10)$$

with

$$p\left(\gamma_t^{(R)} = \tilde{\gamma}_0^{(R)}\right) = \frac{1}{\sqrt{\pi}} e^{-\tilde{\gamma}_0^{(R)\,2}} \qquad (C.11)$$

and with Euler's approximation for the transition probabilities:

$$p\left(\gamma_{t_i}^{(R)} = \tilde{\gamma}_i^{(R)} \mid \gamma_{t_{i-1}}^{(R)} = \tilde{\gamma}_{i-1}^{(R)}\right) = \frac{1}{\sqrt{\pi \mathcal{B}\Delta t}} e^{-\frac{\left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)}(1 - \mathcal{B}\Delta t/2)\right)^2}{\mathcal{B}\Delta t}}. \qquad (C.12)$$

If we take the log of the likelihood function, we have:

$$= \ln\left(\frac{1}{\sqrt{\pi}} e^{-\tilde{\gamma}_0^{(R)\,2}}\right) - \sum_{i=1}^{n} \ln(\sqrt{\pi \mathcal{B}\Delta t}) - \sum_{i=1}^{n} \frac{\left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)}\left(1 - \mathcal{B}\Delta t/2\right)\right)^2}{\mathcal{B}\Delta t}. \qquad (C.13)$$

We now differentiate with respect to $\mathcal{B}$:

$$
\begin{aligned}
\frac{\partial}{\partial \mathcal{B}} \ln \mathcal{L}(\mathcal{B}) &= -\frac{n}{2}\frac{1}{\mathcal{B}} - \frac{\partial}{\partial \mathcal{B}} \sum_{i=1}^{n} \frac{\left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)}(1 - \mathcal{B}\Delta t/2)\right)^2}{\mathcal{B}\Delta t} \\
&= -\frac{n}{2}\frac{1}{\mathcal{B}} - \frac{\partial}{\partial \mathcal{B}} \sum_{i=1}^{n} \frac{\tilde{\gamma}_i^{(R)\,2} - 2\tilde{\gamma}_i^{(R)}\tilde{\gamma}_{i-1}^{(R)}(1 - \mathcal{B}\Delta t/2) + \tilde{\gamma}_{i-1}^{(R)\,2}(1 - \mathcal{B}\Delta t/2)^2}{\mathcal{B}\Delta t} \\
&= -\frac{n}{2}\frac{1}{\mathcal{B}} - \frac{\partial}{\partial \mathcal{B}} \sum_{i=1}^{n} \frac{\tilde{\gamma}_i^{(R)\,2} - 2\tilde{\gamma}_i^{(R)}\tilde{\gamma}_{i-1}^{(R)} + \mathcal{B}\Delta t\, \tilde{\gamma}_i^{(R)}\tilde{\gamma}_{i-1}^{(R)} + \tilde{\gamma}_{i-1}^{(R)\,2}\left(1 - \mathcal{B}\Delta t + \frac{\mathcal{B}^2\Delta t^2}{4}\right)}{\mathcal{B}\Delta t} \\
&= -\frac{n}{2}\frac{1}{\mathcal{B}} - \frac{\partial}{\partial \mathcal{B}} \sum_{i=1}^{n} \left(\frac{\left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)}\right)^2}{\mathcal{B}\Delta t} + \tilde{\gamma}_{i-1}^{(R)}(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)}) + \frac{\tilde{\gamma}_{i-1}^{(R)\,2}}{4}\mathcal{B}\Delta t \right) \\
&= -\frac{n}{2}\frac{1}{\mathcal{B}} + \sum_{i=1}^{n} \frac{\left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)}\right)^2}{\Delta t}\frac{1}{\mathcal{B}^2} - \sum_{i=1}^{n} \frac{\tilde{\gamma}_{i-1}^{(R)\,2}\Delta t}{4} \qquad \text{(C.14)}
\end{aligned}
$$

The optimization condition is:

$$
\frac{\partial}{\partial \mathcal{B}} \ln \mathcal{L}(\mathcal{B}) = 0,
$$
$$
\mathcal{B}^2 \frac{\partial}{\partial \mathcal{B}} \ln \mathcal{L}(\mathcal{B}) = 0, \qquad \text{(C.15)}
$$

which states that:

$$
-\sum_{i=1}^{n} \frac{\tilde{\gamma}_{i-1}^{(R)\,2}\Delta t}{4}\mathcal{B}^2 - \frac{n}{2}\mathcal{B} + \sum_{i=1}^{n} \frac{\left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)}\right)^2}{\Delta t} = 0, \qquad \text{(C.16)}
$$

which is the second line of equation (5.33).

C.2 Proof of equation (5.41)

C.2.1 Proof of the first line of equation (5.41)

We want to prove the first line of equation (5.41), *i.e.* show how Nowman's approximation in ML estimation leads to a third-order polynomial. We remind that the SDE for x_t is:

$$
\mathrm{d}x_t = \mathcal{A}(1 - x_t)\mathrm{d}t + \left(2\frac{\mathcal{A}}{\alpha}x_t\right)^{\frac{1}{2}}\mathrm{d}W_t^{(x)}. \qquad \text{(C.17)}
$$

The generic SDE with linear drift is:

$$
\mathrm{d}X_t = \kappa(\mu - X(t))\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t. \qquad \text{(C.18)}
$$

The SDE of x_t is of this form with $\mu = 1$, $\kappa = \mathcal{A}$ and $\sigma = \left(2\frac{\mathcal{A}}{\alpha}x_t\right)^{\frac{1}{2}}$. According to equation (5.39), we have:

$$
\begin{aligned}
x_{t_i} &= x_{t_{i-1}}e^{-\mathcal{A}\Delta t} + 1 - e^{-\mathcal{A}\Delta t} + \mathcal{N}\left(0, \frac{\frac{2\mathcal{A}x_{t_{i-1}}}{\alpha}(1 - e^{-2\mathcal{A}\Delta t})}{2\mathcal{A}}\right) \\
\Leftrightarrow x_{t_i} &= 1 + e^{-\mathcal{A}\Delta t}(x_{t_{i-1}} - 1) + \mathcal{N}\left(0, \frac{x_{t_{i-1}}(1 - e^{-2\mathcal{A}\Delta t})}{\alpha}\right). \qquad \text{(C.19)}
\end{aligned}
$$

The corresponding transition probabilities are:

$$p(x_{t_i} = \tilde{x}_i \mid x_{t_{i-1}} = \tilde{x}_{i-1}) = \frac{\sqrt{\alpha}}{\sqrt{2\pi\tilde{x}_{i-1}(1 - e^{-2\mathcal{A}\Delta t})}} e^{-\frac{1}{2}\frac{\alpha\left(\tilde{x}_i - 1 + e^{-\mathcal{A}\Delta t}(1 - \tilde{x}_{i-1})\right)^2}{\tilde{x}_{i-1}(1 - e^{-2\mathcal{A}\Delta t})}} \tag{C.20}$$

From equations (C.1), (C.20) and (3.70) for the stationary distribution, we have:

$$\mathcal{L}(\mathcal{A}) = \frac{\alpha^\alpha \tilde{x}_0^{\alpha-1} e^{-\alpha\tilde{x}_0}}{\Gamma(\alpha)} \prod_{i=1}^{n} \frac{\sqrt{\alpha}}{\sqrt{2\pi\tilde{x}_{i-1}(1 - e^{-2\mathcal{A}\Delta t})}} e^{-\frac{1}{2}\frac{\alpha\left(\tilde{x}_i - 1 + e^{-\mathcal{A}\Delta t}(1 - \tilde{x}_{i-1})\right)^2}{\tilde{x}_{i-1}(1 - e^{-2\mathcal{A}\Delta t})}}, \tag{C.21}$$

where we denote $\mathcal{L}(\mathcal{A}) = \mathcal{L}(\mathcal{A}, \alpha)$ since it is assumed that α is known. Taking the log, we get:

$$\begin{aligned}
\ln \mathcal{L}(\mathcal{A}) &= \ln\left(\frac{\alpha^\alpha \tilde{x}_0^{\alpha-1} e^{-\alpha\tilde{x}_0}}{\Gamma(\alpha)}\right) + \sum_{i=1}^{n} \ln\left(\frac{\sqrt{\alpha}}{\sqrt{2\pi\tilde{x}_{i-1}(1 - e^{-2\mathcal{A}\Delta t})}}\right) \\
&\quad + \sum_{i=1}^{n} \frac{\alpha}{2} \frac{\left(\tilde{x}_i - 1 + e^{-\mathcal{A}\Delta t}(1 - \tilde{x}_{i-1})\right)^2}{\tilde{x}_{i-1}(e^{-2\mathcal{A}\Delta t} - 1)} \\
&= \ln\left(\frac{\alpha^\alpha \tilde{x}_0^{\alpha-1} e^{-\alpha\tilde{x}_0}}{\Gamma(\alpha)}\right) - \frac{n}{2}\ln\left(1 - e^{-2\mathcal{A}\Delta t}\right) + \sum_{i=1}^{n} \ln\left(\frac{\sqrt{\alpha}}{\sqrt{2\pi\tilde{x}_{i-1}}}\right) \\
&\quad + \sum_{i=1}^{n} \frac{\alpha}{2} \frac{\left(\tilde{x}_i - 1 + e^{-\mathcal{A}\Delta t}(1 - \tilde{x}_{i-1})\right)^2}{\tilde{x}_{i-1}(e^{-2\mathcal{A}\Delta t} - 1)}. \tag{C.22}
\end{aligned}$$

Since the optimality condition is $\frac{\partial}{\partial\mathcal{A}} \ln \mathcal{L}(\mathcal{A}) = 0$, we now differentiate with respect to $\mathcal{A}$. Some terms readily disappear and we get:

$$\frac{\partial}{\partial\mathcal{A}} \ln \mathcal{L}(\mathcal{A}) = \frac{\partial}{\partial\mathcal{A}}\left(-\frac{n}{2}\ln\left(1 - e^{-2\mathcal{A}\Delta t}\right)\right) + \sum_{i=1}^{n} \frac{\partial}{\partial\mathcal{A}} \frac{\alpha}{2} \frac{\left(\tilde{x}_i - 1 + e^{-\mathcal{A}\Delta t}(1 - \tilde{x}_{i-1})\right)^2}{\tilde{x}_{i-1}(e^{-2\mathcal{A}\Delta t} - 1)}. \tag{C.23}$$

We compute the first term in equation (C.23):

$$\frac{\partial}{\partial\mathcal{A}}\left(-\frac{n}{2}\ln\left(1 - e^{-2\mathcal{A}\Delta t}\right)\right) = -\frac{n}{2}\frac{2\Delta t e^{-2\mathcal{A}\Delta t}}{1 - e^{-2\mathcal{A}\Delta t}} = \frac{n\Delta t e^{-2\mathcal{A}\Delta t}}{e^{-2\mathcal{A}\Delta t} - 1}. \tag{C.24}$$

To derive the terms under the sum in equation (C.23), we set:

$$\begin{aligned}
u &= \left(\tilde{x}_i - 1 + e^{-\mathcal{A}\Delta t}(1 - \tilde{x}_{i-1})\right)^2 \tag{C.25} \\
v &= (e^{-2\mathcal{A}\Delta t} - 1). \tag{C.26}
\end{aligned}$$

If we differentiate u and v with respect to $\mathcal{A}$, we get:

$$\begin{aligned}
u' &= 2\left(\tilde{x}_i - 1 + e^{-\mathcal{A}\Delta t}(1 - \tilde{x}_{i-1})\right)\Delta t e^{-\mathcal{A}\Delta t}(\tilde{x}_{i-1} - 1) \tag{C.27} \\
v' &= -2\Delta t e^{-2\mathcal{A}\Delta t}. \tag{C.28}
\end{aligned}$$

We have now:

$$\begin{aligned}
\frac{\partial}{\partial\mathcal{A}}\frac{u}{v} &= \frac{u'v - v'u}{v^2} \\
&= \frac{2\Delta t\left(\tilde{x}_i - 1 + e^{-\mathcal{A}\Delta t}(1 - \tilde{x}_{i-1})\right)e^{-\mathcal{A}\Delta t}(\tilde{x}_{i-1} - 1)(e^{-2\mathcal{A}\Delta t} - 1)}{(e^{-2\mathcal{A}\Delta t} - 1)^2} \\
&\quad + \frac{2\Delta t e^{-2\mathcal{A}\Delta t}\left(\tilde{x}_i - 1 + e^{-\mathcal{A}\Delta t}(1 - \tilde{x}_{i-1})\right)^2}{(e^{-2\mathcal{A}\Delta t} - 1)^2}. \tag{C.29}
\end{aligned}$$

We now get back to equation (C.23) and multiply it by $(e^{-2\mathcal{A}\Delta t} - 1)^2$. If we express the optimality condition and use the derivatives equations (C.24) and (C.29), we get:

$$
\begin{aligned}
0 &= \frac{\partial}{\partial \mathcal{A}} \ln \mathcal{L}(\mathcal{A}) \\
\Leftrightarrow 0 &= n\Delta t e^{-2\mathcal{A}\Delta t}(e^{-2\mathcal{A}\Delta t} - 1) \quad (C.30) \\
&+ \sum_{i=1}^{n} \frac{\alpha 2\Delta t}{2\tilde{x}_{i-1}} \left(\tilde{x}_i - 1 + e^{-\mathcal{A}\Delta t}(1 - \tilde{x}_{i-1})\right) e^{-\mathcal{A}\Delta t}(\tilde{x}_{i-1} - 1)(e^{-2\mathcal{A}\Delta t} - 1) \\
&+ \sum_{i=1}^{n} \frac{\alpha 2\Delta t}{2\tilde{x}_{i-1}} e^{-2\mathcal{A}\Delta t} \left(\tilde{x}_i - 1 + e^{-\mathcal{A}\Delta t}(1 - \tilde{x}_{i-1})\right)^2. \quad (C.31)
\end{aligned}
$$

We can now simplify this equation by dividing both sides by Δt and by setting $X = e^{-\mathcal{A}\Delta t}$. We get:

$$
\begin{aligned}
0 &= nX^2(X^2 - 1) \quad (C.32) \\
&+ \sum_{i=1}^{n} \frac{\alpha 2\Delta t}{2\tilde{x}_{i-1}} \left[(\tilde{x}_i - 1 + X(1 - \tilde{x}_{i-1}))X(\tilde{x}_{i-1} - 1)(X^2 - 1) + X^2(\tilde{x}_i - 1 + X(1 - \tilde{x}_{i-1}))^2\right].
\end{aligned}
$$

We divide both sides by X since it can be factorized and develop the term under the sum:

$$
\frac{\alpha(\tilde{x}_{i-1} - 1)}{\tilde{x}_{i-1}} \left[X^2(\tilde{x}_i - 1) - (\tilde{x}_i - 1) - X^3(\tilde{x}_{i-1} - 1) + X(\tilde{x}_{i-1} - 1)\right] \quad (C.33)
$$

$$
+ \frac{\alpha}{\tilde{x}_{i-1}} \left[X(\tilde{x}_i - 1)^2 - 2X^2(\tilde{x}_i - 1)(\tilde{x}_{i-1} - 1) + X^3(\tilde{x}_{i-1} - 1)^2\right], \quad (C.34)
$$

which is equal to:

$$
\begin{aligned}
&\frac{\alpha}{\tilde{x}_{i-1}}(X^2(\tilde{x}_i - 1)(\tilde{x}_{i-1} - 1) - (\tilde{x}_i - 1)(\tilde{x}_{i-1} - 1) - X^3(\tilde{x}_{i-1} - 1)^2 + X(\tilde{x}_{i-1} - 1)^2 \\
&+ X(\tilde{x}_i - 1)^2 - 2X^2(\tilde{x}_i - 1)(\tilde{x}_{i-1} - 1) + X^3(\tilde{x}_{i-1} - 1)^2),
\end{aligned}
$$

which after simplification of the X^3 terms yields:

$$
-\frac{\alpha(\tilde{x}_i - 1)(\tilde{x}_{i-1} - 1)}{\tilde{x}_{i-1}}X^2 + \frac{\alpha(\tilde{x}_{i-1} - 1)^2 + \alpha(\tilde{x}_i - 1)^2}{\tilde{x}_{i-1}}X - \frac{\alpha(\tilde{x}_i - 1)(\tilde{x}_{i-1} - 1)}{\tilde{x}_{i-1}}. \quad (C.35)
$$

Combining equations (C.32) and (C.35) (and reminding that we have simplified by X), we obtain:

$$
\begin{aligned}
0 &= nX(X^2 - 1) - \sum_{i=1}^{n} \frac{\alpha(\tilde{x}_i - 1)(\tilde{x}_{i-1} - 1)}{\tilde{x}_{i-1}}X^2 + \sum_{i=1}^{n} \frac{\alpha(\tilde{x}_{i-1} - 1)^2 + \alpha(\tilde{x}_i - 1)^2}{\tilde{x}_{i-1}}X \\
&- \sum_{i=1}^{n} \frac{\alpha(\tilde{x}_i - 1)(\tilde{x}_{i-1} - 1)}{\tilde{x}_{i-1}}, \quad (C.36)
\end{aligned}
$$

which gives the first line of equation (5.41), namely:

$$
\begin{aligned}
nX^3 &- \sum_{i=1}^{n} \frac{\alpha(\tilde{x}_i - 1)(\tilde{x}_{i-1} - 1)}{\tilde{x}_{i-1}}X^2 + \left(-n + \sum_{i=1}^{n} \frac{\alpha(\tilde{x}_i - 1)^2 + \alpha(\tilde{x}_{i-1} - 1)^2}{\tilde{x}_{i-1}}\right)X \\
&- \sum_{i=1}^{n} \frac{\alpha(\tilde{x}_i - 1)(\tilde{x}_{i-1} - 1)}{\tilde{x}_{i-1}} = 0. \quad (C.37)
\end{aligned}
$$

C.2.2   Proof of the second line of equation (5.41)

We now prove the second line of equation (5.41). We remind that the SDE for $\gamma_t^{(R)}$ (or $\gamma_t^{(I)}$, they are identical) is:

$$\mathrm{d}\gamma_t^{(R)} = -\frac{1}{2}\mathcal{B}\gamma_t^{(R)}\mathrm{d}t + \frac{1}{\sqrt{2}}\mathcal{B}^{\frac{1}{2}}\mathrm{d}W_t^{(R)}. \tag{C.38}$$

The SDE of $\gamma_t^{(R)}$ is of the form of equation (C.18) with $\mu = 0$, $\kappa = \frac{\mathcal{B}}{2}$ and $\sigma = \left(\frac{\mathcal{B}}{2}\right)$. According to equation (5.39), we have:

$$\gamma_{t_i}^{(R)} = \gamma_{t_{i-1}}^{(R)}e^{-\mathcal{B}\Delta t/2} + \mathcal{N}\left(0, \frac{1-e^{-\mathcal{B}\Delta t}}{2}\right). \tag{C.39}$$

We obtain the following transition probability:

$$p\left(\gamma_{t_i}^{(R)} = \tilde{\gamma}_i^{(R)} \mid \gamma_{t_{i-1}}^{(R)} = \tilde{\gamma}_{i-1}^{(R)}\right) = \frac{1}{\sqrt{\pi(1-e^{-\mathcal{B}\Delta t})}}e^{-\frac{\left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)}e^{-\mathcal{B}\Delta t/2}\right)^2}{1-e^{-\mathcal{B}\Delta t}}}. \tag{C.40}$$

This is actually the exact transition probability of $\gamma_t^{(R)}$ (see section 5.2.3). From equations (C.10), (C.40) and (3.70) for the stationary distribution, we have:

$$\mathcal{L}(\mathcal{B}) = \frac{1}{\sqrt{\pi}}e^{-\tilde{\gamma}_0^{(R)\,2}}\prod_{i=1}^{n}\frac{1}{\sqrt{\pi(1-e^{-\mathcal{B}\Delta t})}}e^{-\frac{\left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)}e^{-\mathcal{B}\Delta t/2}\right)^2}{1-e^{-\mathcal{B}\Delta t}}}. \tag{C.41}$$

Taking the log, we get:

$$\begin{aligned}
\ln\mathcal{L}(\mathcal{B}) &= \ln\left(\frac{1}{\sqrt{\pi}}e^{-\tilde{\gamma}_0^{(R)\,2}}\right) + n\ln\left(\frac{1}{\sqrt{\pi(1-e^{-\mathcal{B}\Delta t})}}\right) + \sum_{i=1}^{n}\frac{\left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)}e^{-\mathcal{B}\Delta t/2}\right)^2}{e^{-\mathcal{B}\Delta t} - 1} \\
&= \ln\left(\frac{1}{\sqrt{\pi}}e^{-\tilde{\gamma}_0^{(R)\,2}}\right) - \frac{n}{2}\ln\pi - \frac{n}{2}\ln\left(1-e^{-\mathcal{B}\Delta t}\right) \\
&\quad + \sum_{i=1}^{n}\frac{\left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)}e^{-\mathcal{B}\Delta t/2}\right)^2}{e^{-\mathcal{B}\Delta t} - 1}.
\end{aligned} \tag{C.42}$$

We now differentiate with respect to $\mathcal{B}$:

$$\frac{\partial}{\partial\mathcal{B}}\ln\mathcal{L}(\mathcal{B}) = -\frac{n}{2}\frac{\Delta t e^{-\mathcal{B}\Delta t}}{1-e^{-\mathcal{B}\Delta t}} + \sum_{i=1}^{n}\frac{\partial}{\partial\mathcal{B}}\frac{\left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)}e^{-\mathcal{B}\Delta t/2}\right)^2}{e^{-\mathcal{B}\Delta t} - 1}. \tag{C.43}$$

We set i and differentiate the term $\frac{\left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)}e^{-\mathcal{B}\Delta t/2}\right)^2}{e^{-\mathcal{B}\Delta t}-1}$ with respect to $\mathcal{B}$. For convenience, we set:

$$u = \left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)}e^{-\mathcal{B}\Delta t/2}\right)^2 \tag{C.44}$$

$$v = e^{-\mathcal{B}\Delta t} - 1. \tag{C.45}$$

We have:

$$\frac{\partial u}{\partial\mathcal{B}} = 2\left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)}e^{-\mathcal{B}\Delta t/2}\right)\frac{\Delta t}{2}\tilde{\gamma}_{i-1}^{(R)}e^{-\mathcal{B}\Delta t/2} = \Delta t\tilde{\gamma}_{i-1}^{(R)}\left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)}e^{-\mathcal{B}\Delta t/2}\right)e^{-\mathcal{B}\Delta t/2},$$

$$\tag{C.46}$$

and

$$\frac{\partial v}{\partial \mathcal{B}} = -\Delta t e^{-\mathcal{B}\Delta t}. \tag{C.47}$$

Therefore, we have:

$$
\begin{aligned}
\frac{\partial}{\partial \mathcal{B}} \frac{u}{v} &= \frac{u'v - v'u}{v^2} \\
&= \frac{\Delta t \tilde{\gamma}_{i-1}^{(R)} \left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)} e^{-\mathcal{B}\Delta t/2} \right) e^{-\mathcal{B}\Delta t/2} \left(e^{-\mathcal{B}\Delta t} - 1 \right)}{(e^{-\mathcal{B}\Delta t} - 1)^2} \\
&\quad + \frac{\Delta t e^{-\mathcal{B}\Delta t} \left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)} e^{-\mathcal{B}\Delta t/2} \right)^2}{(e^{-\mathcal{B}\Delta t} - 1)^2}.
\end{aligned}
\tag{C.48}
$$

Using equations (C.43) and (C.48), we get:

$$\frac{\partial}{\partial \mathcal{B}} \ln \mathcal{L}(\mathcal{B}) = \frac{n}{2} \frac{\Delta t e^{-\mathcal{B}\Delta t} \left(e^{-\mathcal{B}\Delta t} - 1 \right)}{(e^{-\mathcal{B}\Delta t} - 1)^2} \tag{C.49}$$

$$+ \sum_{i=1}^{n} \frac{\Delta t \tilde{\gamma}_{i-1}^{(R)} \left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)} e^{-\mathcal{B}\Delta t/2} \right) e^{-\mathcal{B}\Delta t/2} \left(e^{-\mathcal{B}\Delta t} - 1 \right) + \Delta t e^{-\mathcal{B}\Delta t} \left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)} e^{-\mathcal{B}\Delta t/2} \right)^2}{(e^{-\mathcal{B}\Delta t} - 1)^2}.$$

We now set $x = e^{-\mathcal{B}\Delta t/2}$ and obtain:

$$\frac{\partial}{\partial \mathcal{B}} \ln \mathcal{L}(\mathcal{B}) = \frac{n}{2} \frac{\Delta t x^2 (x^2 - 1)}{(x^2 - 1)^2} + \sum_{i=1}^{n} \frac{\Delta t \tilde{\gamma}_{i-1}^{(R)} \left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)} x \right) x (x^2 - 1) + \Delta t x^2 \left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)} x \right)^2}{(x^2 - 1)^2}. \tag{C.50}$$

The optimality condition is:

$$\frac{\partial}{\partial \mathcal{B}} \ln \mathcal{L}(\mathcal{B}) = 0 \tag{C.51}$$

$$\Leftrightarrow \quad \frac{n}{2} \frac{\Delta t x^2 (x^2 - 1)}{(x^2 - 1)^2} + \sum_{i=1}^{n} \frac{\Delta t \tilde{\gamma}_{i-1}^{(R)} \left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)} x \right) x (x^2 - 1) + \Delta t x^2 \left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)} x \right)^2}{(x^2 - 1)^2} = 0$$

$$\Leftrightarrow \quad \frac{n}{2} \Delta t x^2 (x^2 - 1) + \sum_{i=1}^{n} \Delta t \tilde{\gamma}_{i-1}^{(R)} \left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)} x \right) x (x^2 - 1) + \Delta t x^2 \left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)} x \right)^2 = 0$$

$$\Leftrightarrow \quad \frac{n}{2} \Delta t (x^3 - x) + \sum_{i=1}^{n} \Delta t \tilde{\gamma}_{i-1}^{(R)} \left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)} x \right) (x^2 - 1) + \Delta t x \left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)} x \right)^2 = 0$$

We now develop the term under the sum:

$$\Delta t \tilde{\gamma}_{i-1}^{(R)} \left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)} x \right) \left(x^2 - 1 \right) + \Delta t x \left(\tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)} x \right)^2 \tag{C.52}$$

$$= \Delta t \tilde{\gamma}_{i-1}^{(R)} \left(\tilde{\gamma}_i^{(R)} x^2 - \tilde{\gamma}_i^{(R)} - \tilde{\gamma}_{i-1}^{(R)} x^3 + \tilde{\gamma}_{i-1}^{(R)} x \right) + \Delta t x \left(\tilde{\gamma}_i^{(R)\,2} - 2\tilde{\gamma}_i^{(R)} \tilde{\gamma}_{i-1}^{(R)} x + \tilde{\gamma}_{i-1}^{(R)\,2} x^2 \right)$$

$$= \Delta t \tilde{\gamma}_{i-1}^{(R)} \tilde{\gamma}_i^{(R)} x^2 - \Delta t \tilde{\gamma}_{i-1}^{(R)} \tilde{\gamma}_i^{(R)} - \Delta t \tilde{\gamma}_{i-1}^{(R)\,2} x^3 + \Delta t \tilde{\gamma}_{i-1}^{(R)\,2} x + \Delta t \tilde{\gamma}_i^{(R)\,2} x - 2\Delta t \tilde{\gamma}_{i-1}^{(R)} \tilde{\gamma}_i^{(R)} x^2$$

$$+ \quad \Delta t \tilde{\gamma}_{i-1}^{(R)\,2} x^3$$

$$= \left(-\Delta t \tilde{\gamma}_{i-1}^{(R)\,2} + \Delta t \tilde{\gamma}_{i-1}^{(R)\,2} \right) x^3 + \left(\Delta t \tilde{\gamma}_{i-1}^{(R)} \tilde{\gamma}_i^{(R)} - 2\Delta t \tilde{\gamma}_{i-1}^{(R)} \tilde{\gamma}_i^{(R)} \right) x^2 + \left(\Delta t \tilde{\gamma}_{i-1}^{(R)\,2} + \Delta t \tilde{\gamma}_i^{(R)\,2} \right) x$$

$$- \quad \Delta t \tilde{\gamma}_{i-1}^{(R)} \tilde{\gamma}_i^{(R)}$$

$$= -\Delta t \tilde{\gamma}_{i-1}^{(R)} \tilde{\gamma}_i^{(R)} x^2 + \Delta t \left(\tilde{\gamma}_i^{(R)\,2} + \tilde{\gamma}_{i-1}^{(R)\,2} \right) - \Delta t \tilde{\gamma}_{i-1}^{(R)} \tilde{\gamma}_i^{(R)}.$$

If we divide by Δt, the optimality condition becomes:

$$\frac{\partial}{\partial \mathcal{B}} \ln \mathcal{L}(\mathcal{B}) = 0$$

$$\Leftrightarrow \quad \frac{n}{2} x^3 - \frac{n}{2} x - \sum_{i=1}^n \tilde{\gamma}_{i-1}^{(R)} \tilde{\gamma}_i^{(R)} x^2 + \sum_{i=1}^n \left(\tilde{\gamma}_i^{(R)\,2} + \tilde{\gamma}_{i-1}^{(R)\,2} \right) x - \sum_{i=1}^n \tilde{\gamma}_{i-1}^{(R)} \tilde{\gamma}_i^{(R)} = 0.$$

Finally, we get the polynomial:

$$\frac{n}{2} x^3 - \sum_{i=1}^n \tilde{\gamma}_{i-1}^{(R)} \tilde{\gamma}_i^{(R)} x^2 + \left(-\frac{n}{2} + \sum_{i=1}^n \left(\tilde{\gamma}_i^{(R)\,2} + \tilde{\gamma}_{i-1}^{(R)\,2} \right) \right) x - \sum_{i=1}^n \tilde{\gamma}_{i-1}^{(R)} \tilde{\gamma}_i^{(R)} = 0. \tag{C.53}$$

We have proved the second line of equation (5.41). Since it is a third order polynomial, it has only on real root $\tilde{\lambda}$. Remembering that $x = e^{-\mathcal{B}\Delta t/2}$, the estimation of $\mathcal{B}$ is:

$$\tilde{\mathcal{B}} = -\frac{2 \ln \tilde{\lambda}}{\Delta t}. \tag{C.54}$$

Appendix D

Complements to chapter 6

D.1 Calculational details for HK scattering

D.1.1 Inverse of $\sigma_{\Psi_c}^{(HK)}\Delta t$

In chapter 6, we used Euler-Maruyama scheme to approximate the transition probabilities over Δt of the process $\begin{bmatrix} x_t & R_t & I_t \end{bmatrix}^\top$. We obtained equation (6.27) in the case of HK scattering, which states that the approximate transition probability $p_{\Psi_c}^{(k)}$ is a multivariate Gaussian distribution with covariance matrix $\sigma_{\Psi_c}^{(HK)}\Delta t$. It is necessary to invert this matrix to compute the transition probabilities. We remind that for any invertible matrix A, it inverse is:

$$A^{-1} = \frac{1}{\det A}\mathrm{com}(A)^\top, \tag{D.1}$$

where $\mathrm{com}(A)$ is the comatrix of A. We have:

$$
\begin{aligned}
\left(\sigma_{\Psi_c}^{(HK)}\Delta t\right)^{-1} &= \Delta t^{-1}(\sigma_{\Psi_c}^{(HK)})^{-1} \\
&= \frac{1}{\Delta t \det \sigma_{\Psi_c}^{(HK)}}\mathrm{com}\left(\sigma_{\Psi_c}^{(HK)}\right)^\top \tag{D.2} \\
&= \frac{2\alpha}{\Delta t \mathcal{A}\mathcal{B}^2 x_t^3}\mathrm{com}\left(\sigma_{\Psi_c}^{(HK)}\right)^\top, \tag{D.3}
\end{aligned}
$$

where we used equation (6.30) for $\det \sigma_{\Psi_c}^{(HK)}$. Since $\sigma_{\Psi_c}^{(HK)}$ is symmetric, so will be $\mathrm{com}\left(\sigma_{\Psi_c}^{(HK)}\right)$ so we have:

$$\mathrm{com}\left(\sigma_{\Psi_c}^{(HK)}\right)^\top = \mathrm{com}\left(\sigma_{\Psi_c}^{(HK)}\right) = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{12} & c_{22} & c_{23} \\ c_{13} & c_{23} & c_{33} \end{bmatrix}. \tag{D.4}$$

Now we must compute the six coefficient c_{11}, c_{22}, c_{33}, c_{12}, c_{13} and c_{23}. We remind that:

$$\sigma_{\Psi_c}^{(HK)} = \begin{bmatrix} \frac{2\mathcal{A}x_t}{\alpha} & \frac{\mathcal{A}(R_t-\Psi_c^{(R)})}{\alpha} & \frac{\mathcal{A}(I_t-\Psi_c^{(I)})}{\alpha} \\ \frac{\mathcal{A}(R_t-\Psi_c^{(R)})}{\alpha} & \frac{\mathcal{A}}{2\alpha}\frac{(R_t-\Psi_c^{(R)})^2}{x_t} + \frac{\mathcal{B}x_t}{2} & \frac{\mathcal{A}}{2\alpha}\frac{(R_t-\Psi_c^{(R)})(I_t-\Psi_c^{(I)})}{x_t} \\ \frac{\mathcal{A}(I_t-\Psi_c^{(I)})}{\alpha} & \frac{\mathcal{A}}{2\alpha}\frac{(R_t-\Psi_c^{(R)})(I_t-\Psi_c^{(I)})}{x_t} & \frac{\mathcal{A}}{2\alpha}\frac{(I_t-\Psi_c^{(I)})^2}{x_t} + \frac{\mathcal{B}x_t}{2} \end{bmatrix}, \tag{D.5}$$

207

from which it is straightforward to compute the c_{ij}. We have:

$$c_{11} = \left[\frac{\mathcal{A}}{2\alpha}\frac{\left(R_t - \Psi_c^{(R)}\right)^2}{x_t} + \frac{\mathcal{B}x_t}{2}\right]\left[\frac{\mathcal{A}}{2\alpha}\frac{\left(I_t - \Psi_c^{(I)}\right)^2}{x_t} + \frac{\mathcal{B}x_t}{2}\right]$$

$$- \left(\frac{\mathcal{A}}{2\alpha}\right)^2\frac{\left(R_t - \Psi_c^{(R)}\right)^2\left(I_t - \Psi_c^{(I)}\right)^2}{x_t^2}$$

$$= \frac{\mathcal{A}}{2\alpha}\frac{\left(R_t - \Psi_c^{(R)}\right)^2}{x_t}\frac{\mathcal{B}x_t}{2} + \frac{\mathcal{A}}{2\alpha}\frac{\left(I_t - \Psi_c^{(I)}\right)^2}{x_t}\frac{\mathcal{B}x_t}{2} + \frac{\mathcal{B}^2 x_t^2}{4}$$

$$\Leftrightarrow c_{11} = \frac{\mathcal{A}\mathcal{B}}{4\alpha}\left(\left(R_t - \Psi_c^{(R)}\right)^2 + \left(I_t - \Psi_c^{(I)}\right)^2\right) + \frac{\mathcal{B}^2 x_t^2}{4}, \tag{D.6}$$

and

$$c_{22} = \frac{2\mathcal{A}x_t}{\alpha}\left[\frac{\mathcal{A}}{2\alpha}\frac{\left(I_t - \Psi_c^{(I)}\right)^2}{x_t} + \frac{\mathcal{B}x_t}{2}\right] - \frac{\mathcal{A}^2\left(I_t - \Psi_c^{(I)}\right)^2}{\alpha^2}$$

$$= \frac{\mathcal{A}^2}{\alpha^2}\left(I_t - \Psi_c^{(I)}\right)^2 + \frac{\mathcal{A}\mathcal{B}x_t^2}{\alpha} - \frac{\mathcal{A}^2}{\alpha^2}\left(I_t - \Psi_c^{(I)}\right)^2$$

$$\Leftrightarrow c_{22} = \frac{\mathcal{A}\mathcal{B}x_t^2}{\alpha}, \tag{D.7}$$

and

$$c_{33} = \frac{2\mathcal{A}x_t}{\alpha}\left[\frac{\mathcal{A}}{2\alpha}\frac{\left(R_t - \Psi_c^{(R)}\right)^2}{x_t} + \frac{\mathcal{B}x_t}{2}\right] - \frac{\mathcal{A}^2\left(R_t - \Psi_c^{(R)}\right)^2}{\alpha^2}$$

$$= \frac{\mathcal{A}^2}{\alpha^2}\left(R_t - \Psi_c^{(R)}\right)^2 + \frac{\mathcal{A}\mathcal{B}x_t^2}{\alpha} - \frac{\mathcal{A}^2}{\alpha^2}\left(R_t - \Psi_c^{(R)}\right)^2$$

$$\Leftrightarrow c_{33} = \frac{\mathcal{A}\mathcal{B}x_t^2}{\alpha}, \tag{D.8}$$

and

$$c_{12} = -\left[\frac{\mathcal{A}(R_t - \Psi_c^{(R)})}{\alpha}\left(\frac{\mathcal{A}}{2\alpha}\frac{(I_t - \Psi_c^{(I)})^2}{x_t} + \frac{\mathcal{B}x_t}{2}\right) - \frac{\mathcal{A}(I_t - \Psi_c^{(I)})}{\alpha}\frac{\mathcal{A}}{2\alpha}\frac{(R_t - \Psi_c^{(R)})(I_t - \Psi_c^{(I)})}{x_t}\right]$$

$$= -\frac{\mathcal{A}\mathcal{B}}{2\alpha}\left(R_t - \Psi_c^{(R)}\right)x_t, \tag{D.9}$$

and

$$c_{13} = \frac{\mathcal{A}(R_t - \Psi_c^{(R)})}{\alpha}\frac{\mathcal{A}}{2\alpha}\frac{(R_t - \Psi_c^{(R)})(I_t - \Psi_c^{(I)})}{x_t} - \frac{\mathcal{A}(I_t - \Psi_c^{(I)})}{\alpha}\left(\frac{\mathcal{A}}{2\alpha}\frac{(R_t - \Psi_c^{(R)})^2}{x_t} + \frac{\mathcal{B}x_t}{2}\right)$$

$$= -\frac{\mathcal{A}\mathcal{B}}{2\alpha}\left(I_t - \Psi_c^{(I)}\right)x_t, \tag{D.10}$$

and

$$c_{23} = -\left[\frac{2\mathcal{A}x_t}{\alpha}\frac{\mathcal{A}}{2\alpha}\frac{(R_t - \Psi_c^{(R)})(I_t - \Psi_c^{(I)})}{x_t} - \frac{\mathcal{A}(I_t - \Psi_c^{(I)})}{\alpha}\frac{\mathcal{A}(R_t - \Psi_c^{(R)})}{\alpha}\right]$$

$$= 0. \tag{D.11}$$

All put together, we obtain the following equation for $\left(\sigma_{\Psi_c}^{(HK)}\Delta t\right)^{-1}$:

$$\left(\sigma_{\Psi_c}^{(HK)}\Delta t\right)^{-1} = \frac{2\alpha}{\Delta t \mathcal{A}\mathcal{B}^2 x_t^3}$$

$$\times \begin{bmatrix} \frac{\mathcal{A}\mathcal{B}}{4\alpha}((R_t - \Psi_c^{(R)})^2 + (I_t - \Psi_c^{(I)})^2) + \frac{\mathcal{B}^2 x_t^2}{4} & -\frac{\mathcal{A}\mathcal{B}}{2\alpha}(R_t - \Psi_c^{(R)})x_t & -\frac{\mathcal{A}\mathcal{B}}{2\alpha}(I_t - \Psi_c^{(I)})x_t \\ -\frac{\mathcal{A}\mathcal{B}}{2\alpha}(R_t - \Psi_c^{(R)})x_t & \frac{\mathcal{A}\mathcal{B}x_t^2}{\alpha} & 0 \\ -\frac{\mathcal{A}\mathcal{B}}{2\alpha}(I_t - \Psi_c^{(I)})x_t & 0 & \frac{\mathcal{A}\mathcal{B}x_t^2}{\alpha} \end{bmatrix},$$

i.e.

$$\left(\sigma_{\Psi_c}^{(HK)}\Delta t\right)^{-1} = \begin{bmatrix} \frac{(R_t - \Psi_c^{(R)})^2 + (I_t - \Psi_c^{(I)})^2}{2\mathcal{B}\Delta t x_t^3} + \frac{\alpha}{2\mathcal{A}\Delta t x_t} & -\frac{R_t - \Psi_c^{(R)}}{\mathcal{B}\Delta t x_t^2} & -\frac{I_t - \Psi_c^{(I)}}{\mathcal{B}\Delta t x_t^2} \\ -\frac{R_t - \Psi_c^{(R)}}{\mathcal{B}\Delta t x_t^2} & \frac{2}{\mathcal{B}\Delta t x_t} & 0 \\ -\frac{I_t - \Psi_c^{(I)}}{\mathcal{B}\Delta t x_t^2} & 0 & \frac{2}{\mathcal{B}\Delta t x_t} \end{bmatrix}. \tag{D.12}$$

which is exactly equation (6.31).

D.1.2 Proof of $\tilde{\Psi}_{c,ML}^{(R)}$ for HK scattering

In this section, we prove equation (6.36) for $\tilde{\Psi}_{c,ML}^{(R)}$, the ML estimator of Ψ_c for HK scattering. We remind from chapter 6 that the approximate transition probabilities for small Δt for HK are:

$$p_{\Psi_c}^{(k)} \approx \frac{1}{(2\pi)^{3/2}\left|\sigma_{\Psi_c}^{(HK)}\Delta t\right|^{1/2}}\exp\left(-\frac{1}{2}(v_k - \mu_k)^T\left(\sigma_{\Psi_c}^{(HK)}\Delta t\right)^{-1}(v_k - \mu_k)\right), \tag{D.13}$$

where

$$p_{\Psi_c}^{(k)} = p_{\Psi_c}\left((x_{t_k}, R_{t_k}, I_{t_k}) = (\tilde{x}_k, \tilde{R}_k, \tilde{I}_k) \mid (x_{t_{k-1}}, R_{t_{k-1}}, I_{t_{k-1}}) = (\tilde{x}_{k-1}, \tilde{R}_{k-1}, \tilde{I}_{k-1})\right), \tag{D.14}$$

and

$$\begin{cases} u_k = \begin{bmatrix} \tilde{x}_{k-1} & \tilde{R}_{k-1} & \tilde{I}_{k-1} \end{bmatrix}^\top \\ v_k = \begin{bmatrix} \tilde{x}_k & \tilde{R}_k & \tilde{I}_k \end{bmatrix}^\top \\ \mu_k = u_k + \beta_{\Psi_c}^{(HK)}\left(\tilde{x}_{k-1}, \tilde{R}_{k-1}, \tilde{I}_{k-1}\right)\Delta t. \end{cases} \tag{D.15}$$

The expression of $\left(\sigma_{\Psi_c}^{(HK)}\Delta t\right)^{-1}$ is now given by equation (D.12). Let:

$$w^{(k)} = v_k - \mu_k = \begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix}^\top. \tag{D.16}$$

We first compute the term inside the exponential. We have:

$$\left(\sigma_{\Psi_c}^{(HK)}\Delta t\right)^{-1} w^{(k)} =$$

$$\begin{bmatrix} w_1\left(\frac{(R_t-\Psi_c^{(R)})^2+(I_t-\Psi_c^{(I)})^2}{2\mathcal{B}\Delta t x_t^3}+\frac{\alpha}{2\mathcal{A}\Delta t x_t}\right)-w_2\frac{R_t-\Psi_c^{(R)}}{\mathcal{B}\Delta t x_t^2}-w_3\frac{I_t-\Psi_c^{(I)}}{\mathcal{B}\Delta t x_t^2} \\ -w_1\frac{R_t-\Psi_c^{(R)}}{\mathcal{B}\Delta t x_t^2}+w_2\frac{2}{\mathcal{B}\Delta t x_t} \\ -w_1\frac{I_t-\Psi_c^{(I)}}{\mathcal{B}\Delta t x_t^2}+w_3\frac{2}{\mathcal{B}\Delta t x_t} \end{bmatrix}, \quad \text{(D.17)}$$

from which we get:

$$
\begin{aligned}
w^{(k)\top}\left(\sigma_{\Psi_c}^{(HK)}\Delta t\right)^{-1} w^{(k)} &= w_1^2\left(\frac{(R_t-\Psi_c^{(R)})^2+(I_t-\Psi_c^{(I)})^2}{2\mathcal{B}\Delta t x_t^3}+\frac{\alpha}{2\mathcal{A}\Delta t x_t}\right)+(w_2^2+w_3^2)\frac{2}{\mathcal{B}\Delta t x_t} \\
&\quad - 2w_1 w_2\frac{R_t-\Psi_c^{(R)}}{\mathcal{B}\Delta t x_t^2}-2w_1 w_3\frac{I_t-\Psi_c^{(I)}}{\mathcal{B}\Delta t x_t^2}. \quad \text{(D.18)}
\end{aligned}
$$

The term inside the exponential is therefore:

$$
\begin{aligned}
-\frac{1}{2}w^{(k)\top}\left(\sigma_{\Psi_c}^{(HK)}\Delta t\right)^{-1} w^{(k)} &= -w_1^2\left(\frac{(R_t-\Psi_c^{(R)})^2+(I_t-\Psi_c^{(I)})^2}{4\mathcal{B}\Delta t x_t^3}+\frac{\alpha}{4\mathcal{A}\Delta t x_t}\right) \\
&\quad + -\frac{(w_2^2+w_3^2)}{\mathcal{B}\Delta t x_t}+w_1 w_2\frac{R_t-\Psi_c^{(R)}}{\mathcal{B}\Delta t x_t^2}+w_1 w_3\frac{I_t-\Psi_c^{(I)}}{\mathcal{B}\Delta t x_t^2} \text{(D.19)}
\end{aligned}
$$

From equation (D.15) and the expression of $\beta_{\Psi_c}^{(HK)}$ which is (equation (6.12)):

$$\beta_{\Psi_c}^{(HK)} = \begin{bmatrix} \mathcal{A}(1-x_t) \\ -\frac{\mathcal{A}+\mathcal{B}}{2}\left(R_t-\Psi_c^{(R)}\right)+\frac{\mathcal{A}\left(R_t-\Psi_c^{(R)}\right)}{2x_t}\left(1-\frac{1}{2\alpha}\right) \\ -\frac{\mathcal{A}+\mathcal{B}}{2}\left(I_t-\Psi_c^{(I)}\right)+\frac{\mathcal{A}\left(I_t-\Psi_c^{(I)}\right)}{2x_t}\left(1-\frac{1}{2\alpha}\right) \end{bmatrix}, \quad \text{(D.20)}$$

we get:

$$w^{(k)} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} \tilde{x}_k - \tilde{x}_{k-1} - \mathcal{A}(1-\tilde{x}_{k-1})\Delta t \\ \tilde{R}_k - \tilde{R}_{k-1}+\frac{\mathcal{A}+\mathcal{B}}{2}(\tilde{R}_{k-1}-\Psi_c^{(R)})-\frac{\mathcal{A}}{2}\frac{(\tilde{R}_{k-1}-\Psi_c^{(R)})}{\tilde{x}_{k-1}}(1-\frac{1}{2\alpha}) \\ \tilde{I}_k - \tilde{I}_{k-1}+\frac{\mathcal{A}+\mathcal{B}}{2}(\tilde{I}_{k-1}-\Psi_c^{(I)})-\frac{\mathcal{A}}{2}\frac{(\tilde{I}_{k-1}-\Psi_c^{(I)})}{\tilde{x}_{k-1}}(1-\frac{1}{2\alpha}) \end{bmatrix} \quad \text{(D.21)}$$

We have, relying on the notations of chapter 6, that the log-likelihood of a time series $(\tilde{x}, \tilde{R}, \tilde{I})$ is:

$$l\left(\tilde{x},\tilde{R},\tilde{I};\Psi_c\right) = \ln\left(\frac{\alpha^\alpha \tilde{x}_0^{\alpha-1} e^{-\alpha\tilde{x}_0}}{\pi\tilde{x}_0\Gamma(\alpha)}\right)-\frac{(\tilde{R}_0-\Psi_c^{(R)})^2+(\tilde{I}_0-\Psi_c^{(I)})^2}{\tilde{x}_0}$$

$$-n\ln\left((2\pi)^{3/2}\left|\sigma_{\Psi_c}^{(HK)}\Delta t\right|^{1/2}\right)+\sum_{k=1}^{n}\Phi_k \quad \text{(D.22)}$$

with

$$\Phi_k = -\frac{1}{2}w^{(k)\top}\left(\sigma_{\Psi_c}^{(HK)}\Delta t\right)^{-1} w^{(k)} \quad \text{(D.23)}$$

To estimate Ψ_c, we use the optimality conditions:

$$\begin{cases} \frac{\partial l}{\partial \Psi_c^{(R)}} = 0 \\ \frac{\partial l}{\partial \Psi_c^{(I)}} = 0. \end{cases} \tag{D.24}$$

By symmetry, we need only to derive the equation for $\tilde{\Psi}_{c,ML}^{(R)}$ and that of $\tilde{\Psi}_{c,ML}^{(I)}$ will follow. It is implicit that equation (D.24) is taken at $(\tilde{\Psi}_{c,ML}^{(R)}, \tilde{\Psi}_{c,ML}^{(I)})$. The idea is now to express explicitely $\frac{\partial l}{\partial \Psi_c^{(R)}} = 0$ as a function of $(\tilde{\Psi}_{c,ML}^{(R)}, \tilde{\Psi}_{c,ML}^{(I)})$. We will soon realize that it depends only on $\tilde{\Psi}_{c,ML}^{(R)}$ and we will invert it to obtain $\tilde{\Psi}_{c,ML}^{(R)}$. From equations (D.24) and (D.22), we get:

$$\frac{\partial l}{\partial \Psi_c^{(R)}} = 0 = 0 - \frac{2\left(\tilde{\Psi}_{c,ML}^{(R)} - \tilde{R}_0\right)}{\tilde{x}_0} + \sum_{k=1}^{n} \frac{\partial \Phi_k}{\Psi_c^{(R)}}. \tag{D.25}$$

We have to compute $\frac{\partial \Phi_k}{\Psi_c^{(R)}}$ with Φ_k given by equation (D.19). We notice first that w_1 does not depend on $\Psi_c^{(R)}$ (nor $\Psi_c^{(I)}$). Second, we have:

$$\begin{aligned} w_2^2 &= \left[\tilde{R}_k - \tilde{R}_{k-1} + \left(\frac{\mathcal{A} + \mathcal{B}}{2} - \frac{\mathcal{A}(1 - \frac{1}{2\alpha})}{2\tilde{x}_{k-1}} \right) \left(\tilde{R}_{k-1} - \Psi_c^{(R)} \right) \right]^2 \\ &= \left(\tilde{R}_k - \tilde{R}_{k-1} + \gamma_k \left(\tilde{R}_{k-1} - \Psi_c^{(R)} \right) \right)^2, \end{aligned} \tag{D.26}$$

where we have set:

$$\gamma_k = \frac{\mathcal{A} + \mathcal{B}}{2} - \frac{\mathcal{A}(1 - \frac{1}{2\alpha})}{2\tilde{x}_{k-1}}. \tag{D.27}$$

Similarly, we have:

$$w_3^2 = \left(\tilde{I}_k - \tilde{I}_{k-1} + \gamma_k \left(\tilde{I}_{k-1} - \Psi_c^{(I)} \right) \right)^2,$$

which does not depend on $\Psi_c^{(R)}$. We get:

$$\begin{aligned} \frac{\partial \Phi_k}{\Psi_c^{(R)}} &= w_1^2 \frac{2\left(\tilde{R}_{k-1} - \Psi_c^{(R)} \right)}{4\mathcal{B}\Delta t \tilde{x}_{k-1}^3} - \frac{1}{\mathcal{B}\Delta t \tilde{x}_{k-1}} \left(\frac{\partial w_2^2}{\partial \Psi_c^{(R)}} + \frac{\partial w_3^2}{\partial \Psi_c^{(R)}} \right) \\ &\quad + w_1 \frac{\partial}{\partial \Psi_c^{(R)}} \left[\left(\tilde{R}_k - \tilde{R}_{k-1} + \gamma_k \left(\tilde{R}_{k-1} - \Psi_c^{(R)} \right) \right) \frac{\left(\tilde{R}_{k-1} - \Psi_c^{(R)} \right)}{\mathcal{B}\Delta t \tilde{x}_{k-1}^2} \right] \\ \Leftrightarrow \frac{\partial \Phi_k}{\Psi_c^{(R)}} &= \frac{w_1^2}{2\mathcal{B}\Delta t \tilde{x}_{k-1}^3} \left(\tilde{R}_{k-1} - \Psi_c^{(R)} \right) - \frac{1}{\mathcal{B}\Delta t \tilde{x}_{k-1}} \left[-2\gamma_k(\tilde{R}_k - \tilde{R}_{k-1}) + 2\gamma_k^2 \left(\Psi_c^{(R)} - \tilde{R}_{k-1} \right) \right] \\ &\quad + \frac{w_1}{\mathcal{B}\Delta t \tilde{x}_{k-1}^2} \left(-(\tilde{R}_k - \tilde{R}_{k-1}) + 2\gamma_k \left(\Psi_c^{(R)} - \tilde{R}_{k-1} \right) \right) \\ \Leftrightarrow \frac{\partial \Phi_k}{\Psi_c^{(R)}} &= \frac{\tilde{\Psi}_{c,ML}^{(R)} - \tilde{R}_{k-1}}{\mathcal{B}\Delta t \tilde{x}_{k-1}} \left(-\frac{w_k^2}{2\tilde{x}_{k-1}^2} - 2\gamma_k^2 + \frac{2\gamma_k w_k}{\tilde{x}_{k-1}} \right) + \frac{\tilde{R}_k - \tilde{R}_{k-1}}{\mathcal{B}\Delta t \tilde{x}_{k-1}} \left(2\gamma_k - \frac{w_k}{\tilde{x}_{k-1}} \right). \end{aligned} \tag{D.28}$$

In the last equation, we denoted $w_k = w_1$. We have:

$$w_k = \tilde{x}_k - \tilde{x}_{k-1} - \mathcal{A}(1 - \tilde{x}_{k-1})\Delta t. \tag{D.29}$$

From equations (D.28) and (D.25), we finally get:

$$\frac{\partial l}{\partial \Psi_c^{(R)}}(\tilde{\Psi}_{c,ML}^{(R)}) = 0$$

$$\Leftrightarrow \quad -\frac{2\left(\tilde{\Psi}_{c,ML}^{(R)} - \tilde{R}_0\right)}{\tilde{x}_0} + \sum_{k=1}^{n} \frac{\tilde{\Psi}_{c,ML}^{(R)} - \tilde{R}_{k-1}}{\mathcal{B}\Delta t \tilde{x}_{k-1}} \left(-\frac{w_k^2}{2\tilde{x}_{k-1}^2} - 2\gamma_k^2 + \frac{2\gamma_k w_k}{\tilde{x}_{k-1}}\right)$$

$$+ \sum_{k=1}^{n} \frac{\tilde{R}_k - \tilde{R}_{k-1}}{\mathcal{B}\Delta t \tilde{x}_{k-1}} \left(2\gamma_k - \frac{w_k}{\tilde{x}_{k-1}}\right) = 0. \tag{D.30}$$

It is exactly equation (6.34) of chapter 6. It is readily invertible and we obtain the expression for the estimator of $\Psi_c^{(R)}$:

$$\tilde{\Psi}_{c,ML}^{(R)} = \frac{-\frac{2\tilde{R}_0}{\tilde{x}_0} + \sum_{k=1}^{n} \frac{\tilde{R}_{k-1}}{\mathcal{B}\Delta t \tilde{x}_{k-1}} \left(-\frac{w_k^2}{2\tilde{x}_{k-1}^2} - 2\gamma_k^2 + \frac{2\gamma_k w_k}{\tilde{x}_{k-1}}\right) - \sum_{k=1}^{n} \frac{\tilde{R}_k - \tilde{R}_{k-1}}{\mathcal{B}\Delta t \tilde{x}_{k-1}} \left(2\gamma_k - \frac{w_k}{\tilde{x}_{k-1}}\right)}{\sum_{k=1}^{n} \frac{1}{\mathcal{B}\Delta t \tilde{x}_{k-1}} \left(-\frac{w_k^2}{2\tilde{x}_{k-1}^2} - 2\gamma_k^2 + \frac{2\gamma_k w_k}{\tilde{x}_{k-1}}\right) - \frac{2}{\tilde{x}_0}}. \tag{D.31}$$

The same procedure holds for $\tilde{\Psi}_{c,ML}^{(I)}$.

D.2   Calculational details for GK scattering

D.2.1   Inverse of $\sigma_{\Psi_c}^{(GK)}\Delta t$

To compute the inverse of $\sigma_{\Psi_c}^{(GK)}\Delta t$, we use again equation (D.1). As in section D.1, we have:

$$\begin{aligned}
\left(\sigma_{\Psi_c}^{(GK)}\Delta t\right)^{-1} &= \Delta t^{-1}(\sigma_{\Psi_c}^{(GK)})^{-1} \\
&= \frac{1}{\Delta t \det \sigma_{\Psi_c}^{(GK)}} \mathrm{com}\left(\sigma_{\Psi_c}^{(GK)}\right)^{\top} \tag{D.32} \\
&= \frac{2\alpha}{\Delta t \mathcal{A}\mathcal{B}^2 x_t^3} \mathrm{com}\left(\sigma_{\Psi_c}^{(GK)}\right)^{\top}, \tag{D.33}
\end{aligned}$$

where we used equation (6.43) for $\det \sigma_{\Psi_c}^{(GK)}$. Since $\sigma_{\Psi_c}^{(GK)}$ is symmetric, so will be $\mathrm{com}\left(\sigma_{\Psi_c}^{(GK)}\right)$ so we have:

$$\mathrm{com}\left(\sigma_{\Psi_c}^{(GK)}\right)^{\top} = \mathrm{com}\left(\sigma_{\Psi_c}^{(GK)}\right) = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{12} & c_{22} & c_{23} \\ c_{13} & c_{23} & c_{33} \end{bmatrix}. \tag{D.34}$$

To compute the coefficients c_{ij}, we remind that:

$$\sigma_{\Psi_c}^{(GK)} = \begin{bmatrix} \frac{2\mathcal{A}x_t}{\alpha} & \frac{\mathcal{A}(R_t + \Psi_c^{(R)}\eta x_t)}{\alpha} & \frac{\mathcal{A}(I_t + \Psi_c^{(I)}\eta x_t)}{\alpha} \\ \frac{\mathcal{A}(R_t + \Psi_c^{(R)}\eta x_t)}{\alpha} & \frac{\mathcal{A}}{2\alpha}\frac{(R_t + \Psi_c^{(R)}\eta x_t)^2}{x_t} + \frac{\mathcal{B}x_t}{2} & \frac{\mathcal{A}}{2\alpha}\frac{(R_t + \Psi_c^{(R)}\eta x_t)(I_t + \Psi_c^{(I)}\eta x_t)}{x_t} \\ \frac{\mathcal{A}(I_t + \Psi_c^{(I)}\eta x_t)}{\alpha} & \frac{\mathcal{A}}{2\alpha}\frac{(R_t + \Psi_c^{(R)}\eta x_t)(I_t + \Psi_c^{(I)}\eta x_t)}{x_t} & \frac{\mathcal{A}}{2\alpha}\frac{(I_t + \Psi_c^{(I)}\eta x_t)^2}{x_t} + \frac{\mathcal{B}x_t}{2} \end{bmatrix}. \tag{D.35}$$

We have:

$$
\begin{aligned}
c_{11} &= \left(\frac{\mathcal{A}}{2\alpha} \frac{\left(R_t + \Psi_c^{(R)}\eta x_t\right)^2}{x_t} + \frac{\mathcal{B}x_t}{2} \right) \left(\frac{\mathcal{A}}{2\alpha} \frac{\left(I_t + \Psi_c^{(I)}\eta x_t\right)^2}{x_t} + \frac{\mathcal{B}x_t}{2} \right) \\
&\quad - \left(\frac{\mathcal{A}}{2\alpha} \right)^2 \frac{\left(R_t + \Psi_c^{(R)}\eta x_t\right)^2 \left(I_t + \Psi_c^{(I)}\eta x_t\right)^2}{x_t^2} \\
&= \frac{\mathcal{A}\mathcal{B}}{4\alpha} \left(\left(R_t + \Psi_c^{(R)}\eta x_t\right)^2 + \left(I_t + \Psi_c^{(I)}\eta x_t\right)^2 \right) + \frac{\mathcal{B}x_t^2}{4},
\end{aligned}
\tag{D.36}
$$

and

$$
\begin{aligned}
c_{22} &= \frac{2\mathcal{A}x_t}{\alpha} \left(\frac{\mathcal{A}}{2\alpha} \frac{\left(I_t + \Psi_c^{(I)}\eta x_t\right)^2}{x_t} + \frac{\mathcal{B}x_t}{2} \right) - \frac{\mathcal{A}^2 \left(I_t + \Psi_c^{(I)}\eta x_t\right)^2}{\alpha^2} \\
&= \frac{\mathcal{A}\mathcal{B}x_t^2}{\alpha},
\end{aligned}
\tag{D.37}
$$

and

$$
\begin{aligned}
c_{33} &= \frac{2\mathcal{A}x_t}{\alpha} \left(\frac{\mathcal{A}}{2\alpha} \frac{\left(R_t + \Psi_c^{(R)}\eta x_t\right)^2}{x_t} + \frac{\mathcal{B}x_t}{2} \right) - \frac{\mathcal{A}^2 \left(R_t + \Psi_c^{(R)}\eta x_t\right)^2}{\alpha^2} \\
&= \frac{\mathcal{A}\mathcal{B}x_t^2}{\alpha},
\end{aligned}
\tag{D.38}
$$

and

$$
\begin{aligned}
c_{12} &= - \left[\frac{\mathcal{A}\left(R_t + \Psi_c^{(R)}\eta x_t\right)}{\alpha} \left(\frac{\mathcal{A}}{2\alpha} \frac{\left(I_t + \Psi_c^{(I)}\eta x_t\right)^2}{x_t} + \frac{\mathcal{B}x_t}{2} \right) \right. \\
&\quad \left. - \frac{\mathcal{A}\left(I_t + \Psi_c^{(I)}\eta x_t\right)}{\alpha} \frac{\mathcal{A}}{2\alpha} \frac{\left(R_t + \Psi_c^{(R)}\eta x_t\right)\left(I_t + \Psi_c^{(I)}\eta x_t\right)}{x_t} \right] \\
&= - \left[\frac{\mathcal{A}^2 \left(R_t + \Psi_c^{(R)}\eta x_t\right)\left(I_t + \Psi_c^{(I)}\eta x_t\right)^2}{2\alpha^2} + \frac{\mathcal{A}\mathcal{B}x_t}{2\alpha}\left(R_t + \Psi_c^{(R)}\eta x_t\right) \right. \\
&\quad \left. - \frac{\mathcal{A}^2 \left(R_t + \Psi_c^{(R)}\eta x_t\right)\left(I_t + \Psi_c^{(I)}\eta x_t\right)^2}{2\alpha^2} \right] \\
\Leftrightarrow c_{12} &= - \frac{\mathcal{A}\mathcal{B}x_t}{2\alpha}\left(R_t + \Psi_c^{(R)}\eta x_t\right),
\end{aligned}
\tag{D.40}
$$

(D.39)

and

$$
\begin{aligned}
c_{13} &= \frac{\mathcal{A}\left(R_t + \Psi_c^{(R)}\eta x_t\right)}{\alpha} \frac{\mathcal{A}}{2\alpha} \frac{\left(I_t + \Psi_c^{(I)}\eta x_t\right)\left(R_t + \Psi_c^{(R)}\eta x_t\right)}{x_t} \\
&\quad - \frac{\mathcal{A}\left(I_t + \Psi_c^{(I)}\eta x_t\right)}{\alpha}\left(\frac{\mathcal{A}}{2\alpha}\frac{\left(R_t + \Psi_c^{(R)}\eta x_t\right)^2}{x_t} + \frac{\mathcal{B}x_t}{2}\right) \\
&= \frac{\mathcal{A}^2\left(R_t + \Psi_c^{(R)}\eta x_t\right)^2\left(I_t + \Psi_c^{(I)}\eta x_t\right)}{2\alpha^2} - \frac{\mathcal{A}\mathcal{B}x_t}{2\alpha}\left(I_t + \Psi_c^{(I)}\eta x_t\right) \\
&\quad - \frac{\mathcal{A}^2\left(R_t + \Psi_c^{(R)}\eta x_t\right)^2\left(I_t + \Psi_c^{(I)}\eta x_t\right)}{2\alpha^2} \\
\Leftrightarrow c_{13} &= -\frac{\mathcal{A}\mathcal{B}x_t}{2\alpha}\left(I_t + \Psi_c^{(I)}\eta x_t\right), \tag{D.41}
\end{aligned}
$$

and

$$
\begin{aligned}
c_{23} &= -\left[\frac{2\mathcal{A}x_t}{\alpha}\frac{\mathcal{A}}{2\alpha}\frac{(R_t + \Psi_c^{(R)}\eta x_t)(I_t + \Psi_c^{(I)}\eta x_t)}{x_t} - \frac{\mathcal{A}(I_t + \Psi_c^{(I)}\eta x_t)}{\alpha}\frac{\mathcal{A}(R_t + \Psi_c^{(R)}\eta x_t)}{\alpha}\right] \\
&= 0. \tag{D.42}
\end{aligned}
$$

All put together, we obtain the following equation for $\left(\sigma_{\Psi_c}^{(GK)}\Delta t\right)^{-1}$:

$$
\left(\sigma_{\Psi_c}^{(GK)}\Delta t\right)^{-1} = \frac{2\alpha}{\Delta t \mathcal{A}\mathcal{B}^2 x_t^3}
$$
$$
\times \begin{bmatrix}
\frac{\mathcal{A}\mathcal{B}((R_t + \Psi_c^{(R)}\eta x_t)^2 + (I_t + \Psi_c^{(I)}\eta x_t)^2)}{4\alpha} + \frac{\mathcal{B}x_t^2}{4} & -\frac{\mathcal{A}\mathcal{B}x_t(R_t + \Psi_c^{(R)}\eta x_t)}{2\alpha} & -\frac{\mathcal{A}\mathcal{B}x_t(I_t + \Psi_c^{(I)}\eta x_t)}{2\alpha} \\
-\frac{\mathcal{A}\mathcal{B}x_t(R_t + \Psi_c^{(R)}\eta x_t)}{2\alpha} & \frac{\mathcal{A}\mathcal{B}x_t^2}{\alpha} & 0 \\
-\frac{\mathcal{A}\mathcal{B}x_t(I_t + \Psi_c^{(I)}\eta x_t)}{2\alpha} & 0 & \frac{\mathcal{A}\mathcal{B}x_t^2}{\alpha}
\end{bmatrix},
$$

i.e.

$$
\left(\sigma_{\Psi_c}^{(GK)}\Delta_t\right)^{-1} = \begin{bmatrix}
\frac{(R_t + \Psi_c^{(R)}\eta x_t)^2 + (I_t + \Psi_c^{(I)}\eta x_t)^2}{2\mathcal{B}\Delta t x_t^3} + \frac{\alpha}{2\mathcal{A}\Delta t x_t} & -\frac{R_t + \Psi_c^{(R)}\eta x_t}{\mathcal{B}\Delta t x_t^2} & -\frac{I_t + \Psi_c^{(I)}\eta x_t}{\mathcal{B}\Delta t x_t^2} \\
-\frac{R_t + \Psi_c^{(R)}\eta x_t}{\mathcal{B}\Delta t x_t^2} & \frac{2}{\mathcal{B}\Delta t x_t} & 0 \\
-\frac{I_t + \Psi_c^{(I)}\eta x_t}{\mathcal{B}\Delta t x_t^2} & 0 & \frac{2}{\mathcal{B}\Delta t x_t}
\end{bmatrix}. \tag{D.43}
$$

which is exactly equation (6.44).

D.2.2 Proof of $\tilde{\Psi}_{c,ML}^{(R)}$ for GK scattering

In this section, we prove equation (6.50) for $\tilde{\Psi}_{c,ML}^{(R)}$, the ML estimator of Ψ_c for GK scattering. We remind from chapter 6 that the approximate transition probabilities for small Δt for GK are:

$$
p_{\Psi_c}^{(k)} \approx \frac{1}{(2\pi)^{3/2}\left|\sigma_{\Psi_c}^{(GK)}\Delta t\right|^{1/2}} \exp\left(-\frac{1}{2}(v_k - \mu_k)^T\left(\sigma_{\Psi_c}^{(GK)}\Delta t\right)^{-1}(v_k - \mu_k)\right), \tag{D.44}
$$

where

$$p_{\Psi_c}^{(k)} = p_{\Psi_c}\left((x_{t_k}, R_{t_k}, I_{t_k}) = (\tilde{x}_k, \tilde{R}_k, \tilde{I}_k) \mid (x_{t_{k-1}}, R_{t_{k-1}}, I_{t_{k-1}}) = (\tilde{x}_{k-1}, \tilde{R}_{k-1}, \tilde{I}_{k-1})\right), \quad \text{(D.45)}$$

and

$$\begin{cases} u_k = \begin{bmatrix} \tilde{x}_{k-1} & \tilde{R}_{k-1} & \tilde{I}_{k-1} \end{bmatrix}^\top \\ v_k = \begin{bmatrix} \tilde{x}_k & \tilde{R}_k & \tilde{I}_k \end{bmatrix}^\top \\ \mu_k = u_k + \beta_{\Psi_c}^{(GK)}\left(\tilde{x}_{k-1}, \tilde{R}_{k-1}, \tilde{I}_{k-1}\right)\Delta t. \end{cases} \quad \text{(D.46)}$$

The expression of $\left(\sigma_{\Psi_c}^{(GK)}\Delta t\right)^{-1}$ is now given by equation (D.43). Let:

$$w^{(k)} = v_k - \mu_k = \begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix}^\top. \quad \text{(D.47)}$$

We first compute the term inside the exponential. We notice that $\left(\sigma_{\Psi_c}^{(GK)}\Delta t\right)^{-1}$ (equation (D.43)) has the same form as $\left(\sigma_{\Psi_c}^{(HK)}\Delta t\right)^{-1}$ (equation (D.12)) with $R_t + \Psi_c^{(R)}\eta x_t$ replacing $R_t - \Psi_c^{(R)}$ and $I_t + \Psi_c^{(I)}\eta x_t$ replacing $I_t - \Psi_c^{(I)}$. Therefore:

The term inside the exponential is therefore:

$$-\frac{1}{2}w^{(k)\top}\left(\sigma_{\Psi_c}^{(GK)}\Delta t\right)^{-1}w^{(k)} = -w_1^2\left(\frac{(R_t + \Psi_c^{(R)}\eta x_t)^2 + (I_t + \Psi_c^{(I)}\eta x_t)^2}{4\mathcal{B}\Delta t x_t^3} + \frac{\alpha}{4\mathcal{A}\Delta t x_t}\right)$$

$$+ \quad -\frac{(w_2^2 + w_3^2)}{\mathcal{B}\Delta t x_t} + w_1 w_2 \frac{R_t + \Psi_c^{(R)}\eta x_t}{\mathcal{B}\Delta t x_t^2} + w_1 w_3 \frac{I_t + \Psi_c^{(I)}\eta x_t}{\mathcal{B}\Delta t x_t^2}. \quad \text{(D.48)}$$

From equation (D.46) and the expression of $\beta_{\Psi_c}^{(GK)}$ which is (equation (6.17)):

$$\beta_{\Psi_c}^{(GK)} = \begin{bmatrix} \mathcal{A}(1 - x_t) \\ \eta\Psi_c^{(R)}\mathcal{A}(1 - x_t) + \left(R_t - \Psi_c^{(R)}\eta x_t\right)\left(-\frac{\mathcal{A}+\mathcal{B}}{2} + \frac{\mathcal{A}}{2x_t}\left(1 - \frac{1}{2\alpha}\right)\right) \\ \eta\Psi_c^{(I)}\mathcal{A}(1 - x_t) + \left(I_t - \Psi_c^{(I)}\eta x_t\right)\left(-\frac{\mathcal{A}+\mathcal{B}}{2} + \frac{\mathcal{A}}{2x_t}\left(1 - \frac{1}{2\alpha}\right)\right) \end{bmatrix}, \quad \text{(D.49)}$$

we get:

$$w^{(k)} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \quad \text{(D.50)}$$

$$= \begin{bmatrix} \tilde{x}_k - \tilde{x}_{k-1} - \mathcal{A}(1 - \tilde{x}_{k-1})\Delta t \\ \tilde{R}_k - \tilde{R}_{k-1} - \eta\Psi_c^{(R)}\mathcal{A}(1 - \tilde{x}_{k-1})\Delta t - \left(\tilde{R}_{k-1} - \Psi_c^{(R)}\eta\tilde{x}_{k-1}\right)\left(-\frac{\mathcal{A}+\mathcal{B}}{2} + \frac{\mathcal{A}}{2\tilde{x}_{k-1}}\left(1 - \frac{1}{2\alpha}\right)\right)\Delta t \\ \tilde{I}_k - \tilde{I}_{k-1} - \eta\Psi_c^{(I)}\mathcal{A}(1 - \tilde{x}_{k-1})\Delta t - \left(\tilde{I}_{k-1} - \Psi_c^{(I)}\eta\tilde{x}_{k-1}\right)\left(-\frac{\mathcal{A}+\mathcal{B}}{2} + \frac{\mathcal{A}}{2\tilde{x}_{k-1}}\left(1 - \frac{1}{2\alpha}\right)\right)\Delta t \end{bmatrix}.$$

We have, relying on the notations of chapter 6, that the log-likelihood of a time series $(\tilde{x}, \tilde{R}, \tilde{I})$ is:

$$l\left(\tilde{x}, \tilde{R}, \tilde{I}; \Psi_c\right) = \ln\left(\frac{\alpha^\alpha \tilde{x}_0^{\alpha-1} e^{-\alpha\tilde{x}_0}}{\pi\tilde{x}_0\Gamma(\alpha)}\right) - \frac{(\tilde{R}_0 - \Psi_c^{(R)}\eta\tilde{x}_0)^2 + (\tilde{I}_0 - \Psi_c^{(I)}\eta\tilde{x}_0)^2}{\tilde{x}_0}$$

$$-n\ln\left((2\pi)^{3/2}\left|\sigma_{\Psi_c}^{(GK)}\Delta t\right|^{1/2}\right) + \sum_{k=1}^{n}\Phi_k \quad \text{(D.51)}$$

with
$$\Phi_k = -\frac{1}{2} w^{(k)\top} \left(\sigma_{\Psi_c}^{(GK)} \Delta t \right)^{-1} w^{(k)} \tag{D.52}$$
To estimate Ψ_c, we use the optimality conditions:
$$\begin{cases} \frac{\partial l}{\partial \Psi_c^{(R)}} = 0 \\ \frac{\partial l}{\partial \Psi_c^{(I)}} = 0. \end{cases} \tag{D.53}$$

By symmetry, we need only to derive the equation for $\tilde{\Psi}_{c,ML}^{(R)}$ and that of $\tilde{\Psi}_{c,ML}^{(I)}$ will follow. It is again implicit that equation (D.53) is taken at $(\tilde{\Psi}_{c,ML}^{(R)}, \tilde{\Psi}_{c,ML}^{(I)})$. The idea is now to express explicitely $\frac{\partial l}{\partial \Psi_c^{(R)}} = 0$ as a function of $(\tilde{\Psi}_{c,ML}^{(R)}, \tilde{\Psi}_{c,ML}^{(I)})$. As for HK scattering, it will only depende on $\tilde{\Psi}_{c,ML}^{(R)}$ and we will invert it to obtain $\tilde{\Psi}_{c,ML}^{(R)}$. From equations (D.53) and (D.51), we get:

$$\frac{\partial l}{\partial \Psi_c^{(R)}} = 0 = 0 - \frac{2\eta \tilde{x}_0 \left(\tilde{\Psi}_{c,ML}^{(R)} \eta \tilde{x}_0 - \tilde{R}_0 \right)}{\tilde{x}_0} + \sum_{k=1}^{n} \frac{\partial \Phi_k}{\Psi_c^{(R)}}. \tag{D.54}$$

We have to compute $\frac{\partial \Phi_k}{\Psi_c^{(R)}}$ with Φ_k given by equation (D.48). We notice that w_1 still does not depend on $\Psi_c^{(R)}$ (nor $\Psi_c^{(I)}$). Since w_3 also does not depend on $\Psi_c^{(R)}$, we have:

$$\begin{aligned} \frac{\partial \Phi_k}{\Psi_c^{(R)}} &= -w_1^2 \frac{2\eta \tilde{x}_{k-1} \left(\tilde{R}_{k-1} + \Psi_c^{(R)} \eta \tilde{x}_{k-1} \right)}{4\mathcal{B}\Delta t \tilde{x}_{k-1}^3} - \frac{1}{\mathcal{B}\Delta t \tilde{x}_{k-1}} \frac{\partial w_2^2}{\partial \Psi_c^{(R)}} \\ &+ \frac{w_1}{\mathcal{B}\Delta t \tilde{x}_{k-1}^2} \frac{\partial}{\partial \Psi_c^{(R)}} \left[(\tilde{R}_{k-1} + \Psi_c^{(R)} \eta \tilde{x}_{k-1}) w_2 \right]. \end{aligned}$$

If we set again:
$$\gamma_k = \frac{\mathcal{A} + \mathcal{B}}{2} - \frac{\mathcal{A}(1 - \frac{1}{2\alpha})}{2\tilde{x}_{k-1}}, \tag{D.55}$$
we have from equation (D.50):
$$w_2^2 = \left(\tilde{R}_k - \tilde{R}_{k-1} - \eta \Psi_c^{(R)} \mathcal{A}(1 - \tilde{x}_{k-1})\Delta t - \gamma_k \Delta t \left(\tilde{R}_{k-1} - \Psi_c^{(R)} \eta \tilde{x}_{k-1} \right) \right)^2. \tag{D.56}$$
Therefore on one hand we have:
$$\begin{aligned} \frac{\partial w_2^2}{\partial \Psi_c^{(R)}} &= 2 \left(\tilde{R}_k - \tilde{R}_{k-1} - \eta \Psi_c^{(R)} \mathcal{A}(1 - \tilde{x}_{k-1})\Delta t - \gamma_k \Delta t \left(\tilde{R}_{k-1} - \Psi_c^{(R)} \eta \tilde{x}_{k-1} \right) \right) \\ &\times (-\eta \mathcal{A}(1 - \tilde{x}_{k-1})\Delta t + \gamma_k \Delta t \eta \tilde{x}_{k-1}) \\ &= 2 \left(\tilde{R}_k - \tilde{R}_{k-1} - \gamma_k \Delta t \tilde{R}_{k-1} \right) (-\eta \mathcal{A}(1 - \tilde{x}_{k-1})\Delta t + \gamma_k \Delta t \eta \tilde{x}_{k-1}) \\ &+ 2 (-\eta \mathcal{A}\Delta t(1 - \tilde{x}_{k-1}) + \gamma_k \Delta \eta \tilde{x}_{k-1}) (-\eta \mathcal{A}(1 - \tilde{x}_{k-1})\Delta t + \gamma_k \Delta t \eta \tilde{x}_{k-1}) \Psi_c^{(R)}. \end{aligned} \tag{D.57}$$
On the other hand,
$$\begin{aligned} \frac{\partial}{\partial \Psi_c^{(R)}} & \left[(\tilde{R}_{k-1} + \Psi_c^{(R)} \eta \tilde{x}_{k-1}) w_2 \right] = \eta \tilde{x}_{k-1} \left(\tilde{R}_k - \tilde{R}_{k-1} - \eta \Psi_c^{(R)} \mathcal{A}(1 - \tilde{x}_{k-1})\Delta t \right. \\ &- \gamma_k \Delta t \left(\tilde{R}_{k-1} - \Psi_c^{(R)} \eta \tilde{x}_{k-1} \right) \bigg) + \left(\tilde{R}_{k-1} + \Psi_c^{(R)} \eta \tilde{x}_{k-1} \right) (-\eta \mathcal{A}(1 - \tilde{x}_{k-1})\Delta t + \gamma_k \Delta t \eta \tilde{x}_{k-1}) \\ &= \eta \tilde{x}_{k-1} \left(\tilde{R}_k - \tilde{R}_{k-1} - \gamma_k \Delta t \tilde{R}_{k-1} \right) + \eta \tilde{x}_{k-1} (-\eta \mathcal{A}(1 - \tilde{x}_{k-1})\Delta t + \gamma_k \eta \tilde{x}_{k-1}\Delta t) \Psi_c^{(R)} \\ &+ \tilde{R}_{k-1} (-\eta \mathcal{A}(1 - \tilde{x}_{k-1})\Delta t + \gamma_k \Delta t \eta \tilde{x}_{k-1}) \\ &+ \eta \tilde{x}_{k-1} (-\eta \mathcal{A}(1 - \tilde{x}_{k-1})\Delta t + \gamma_k \Delta t \eta \tilde{x}_{k-1}) \Psi_c^{(R)}. \end{aligned} \tag{D.58}$$

We obtain:

$$\frac{\partial}{\partial \Psi_c^{(R)}} \left[(\tilde{R}_{k-1} + \Psi_c^{(R)} \eta \tilde{x}_{k-1}) w_2 \right] = \eta \tilde{x}_{k-1} \left(\tilde{R}_k - \tilde{R}_{k-1} - \gamma_k \Delta t \tilde{R}_{k-1} \right)$$
$$+ \quad \tilde{R}_{k-1} \left(-\eta \mathcal{A} (1 - \tilde{x}_{k-1}) \Delta t + \gamma_k \Delta t \eta \tilde{x}_{k-1} \right)$$
$$+ \quad 2\eta \tilde{x}_{k-1} \left(-\eta \mathcal{A} (1 - \tilde{x}_{k-1}) \Delta t + \gamma_k \Delta t \eta \tilde{x}_{k-1} \right) \Psi_c^{(R)} \tag{D.59}$$

If we denote $w_k = w_1$ and use equations (D.54), (D.55), (D.57) and (D.59), we finally get:

$$\frac{\partial l}{\partial \Psi_c^{(R)}} (\tilde{\Psi}_{c,ML}^{(R)}, \tilde{\Psi}_{c,ML}^{(I)}) = 0 \tag{D.60}$$

$$\Leftrightarrow \quad -\frac{2\eta \left(\tilde{\Psi}_{c,ML}^{(R)} \eta \tilde{x}_0 - \tilde{R}_0 \right)}{\tilde{x}_0} + \sum_{k=1}^{n} \lambda_k^{(1)} + \Psi_c^{(R)} \sum_{k=1}^{n} \lambda_k^{(2)} = 0 \tag{D.61}$$

with

$$\lambda_k^{(1)} = \frac{-w_k^2 \eta \tilde{R}_{k-1}}{2 \mathcal{B} \Delta t \tilde{x}_{k-1}^2} - \frac{2}{\mathcal{B} \Delta t \tilde{x}_{k-1}} \left(\tilde{R}_k - \tilde{R}_{k-1} - \gamma_k \Delta t \tilde{R}_{k-1} \right)$$
$$\times \left(-\eta \mathcal{A} \Delta t (1 - \tilde{x}_{k-1}) + \gamma_k \Delta t \eta \tilde{x}_{k-1} \right) + \frac{w_k}{\mathcal{B} \Delta t \tilde{x}_{k-1}^2} \eta \tilde{x}_{k-1} \left(\tilde{R}_k - \tilde{R}_{k-1} - \gamma_k \Delta t \tilde{R}_{k-1} \right)$$
$$+ \frac{w_k}{\mathcal{B} \Delta t \tilde{x}_{k-1}^2} \tilde{R}_{k-1} \left(-\eta \mathcal{A} \Delta t (1 - \tilde{x}_{k-1}) + \gamma_k \Delta t \eta \tilde{x}_{k-1} \right) \tag{D.62}$$

and

$$\lambda_k^{(2)} = \frac{-w_k^2 \eta^2}{2 \mathcal{B} \Delta t \tilde{x}_{k-1}} - \frac{2}{\mathcal{B} \Delta t \tilde{x}_{k-1}} \left(-\eta \mathcal{A} \Delta t (1 - \tilde{x}_{k-1}) + \gamma_k \Delta t \eta \tilde{x}_{k-1} \right)$$
$$\times \left(-\eta \mathcal{A} \Delta t (1 - \tilde{x}_{k-1}) + \gamma_k \Delta t \eta \tilde{x}_{k-1} \right)$$
$$+ \frac{2 w_k \eta}{\mathcal{B} \Delta t \tilde{x}_{k-1}} \left(-\eta \mathcal{A} \Delta t (1 - \tilde{x}_{k-1}) + \gamma_k \Delta t \eta \tilde{x}_{k-1} \right) \tag{D.63}$$

which are exactly equations (6.47), (6.48) and (6.49) of chapter 6. It is invertible and we obtain the expression for the estimator of $\Psi_c^{(R)}$:

$$\tilde{\Psi}_{c,ML}^{(R)} = \frac{-\frac{2\eta \tilde{R}_0}{\tilde{x}_0} - \sum_{k=1}^{n} \lambda_k^{(1)}}{-2\eta^2 + \sum_{k=1}^{n} \lambda_k^{(2)}}. \tag{D.64}$$

The same procedure holds for $\tilde{\Psi}_{c,ML}^{(I)}$.

Bibliography

[1] Wikipedia.

[2] http://soma.mcmaster.ca/ipix.php.

[3] https://seashepherd.org/.

[4] https://www.theguardian.com/environment/2019/apr/29/whale-with-harness-could-be-russian-weapon-say-norwegian-experts.

[5] *Small target detection within sea clutter based on fractal analysis*, 2016.

[6] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions*. National Bureau of Standards Applied Mathematics Series, 1964.

[7] Y. Aït-Sahalia. Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica*, 70(1):223–262, 01 2002.

[8] Y. Aït-Sahalia and J. Jacod. *High-Frequency Financial Econometrics*. Princeton University Press, 2014.

[9] M. Ben Alaya and A. Kebaier. Parameter Estimation for the Square-Root Diffusions: Ergodic and Nonergodic Cases. *Stochastic Models*, 28(4):609–634, 2012.

[10] M. Ben Alaya and A. Kebaier. Asymptotic Behavior of the Maximum Likelihood Estimator for Ergodic and Nonergodic Square-Root Diffusions. *Stochastic Analysis and Applications*, 31(4):552–573, 2013.

[11] W. R. Alpers and C. Bruening. On the Relative Importance of Motion-Related Contributions to the SAR Imaging Mechanism of Ocean Surface Waves. *IEEE Transactions on Geoscience and Remote Sensing*, 1986.

[12] R. Altmeyer. Stable Limit Theorems For Estimators of Integrated Volatility. Master's thesis, Freie Universität Berlin, 2013.

[13] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association*, 96(453):42–55, 2001.

[14] J. W. Antony, K. Schmidt, M. Schwerdt, D. Polimeni, N. Tous-Ramon, M. Bachmann, and G. C. Alfonzo. Radiometric Accuracy and Stability of TerraSAR-X and TanDEM-X. In *11th European Conference on Synthetic Aperture Radar*, 2016.

[15] G. B. Arfken and H. J. Weber. *Mathematical Methods for Physicists*. Elsevier Academic Press, 2005.

[16] A. Arnold-Bos. *La surveillance maritime en imagerie radar bistatique: théorie, simulation, contribution à la détection automatique du sillage des navires.* PhD thesis, Université de Bretagne Occidentale, 2010.

[17] M. Y. Ayari, A. Khenchaf, and A. Coatanhay. Simulations of the bistatic scattering using two-scale model and the unified sea spectrum. *Journal of Applied Remote Sensing,* 2007.

[18] A. Balleri, A. Nehorai, and J. Wang. Maximum likelihood estimation for compound-gaussian clutter with inverse gamma texture. *IEEE Transactions on Aerospace and Electronic Systems,* 43(2):775–779, April 2007.

[19] Richard Bamler and Birgit Schättler. SAR Data Acquisition and Image Formation, 1993.

[20] Richard Barakat. Weak-scatterer generalization of the k-density function with application to laser scattering in atmospheric turbulence. *J. Opt. Soc. Am. A,* 3(4):401–409, Apr 1986.

[21] O. E. Barndorff-Nielsen and N. Shephard. Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology),* 64(2):253–280, 2002.

[22] A. Benetazzo, F. Serafino, F. Bergamasco, G. Ludeno, F. Ardhuin, P. Sutherland, M. Sclavo, and F. Barbariol. Stereo imaging and X-band radar wave data fusion: An assessment. *Ocean Engineering,* 152:346 – 352, 2018.

[23] F. Berizzi and E. Dalle Mese. Fractal Analysis of the Signal Scattered from the Sea Surface. *IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION,* 47(2):324–338, 1999.

[24] F. Berizzi and E. Dalle Mese. Sea-wave fractal spectrum for SAR remote sensing. *IEE Proceedings - Radar, Sonar and Navigation,* 148(2):56 – 66, 2001.

[25] E. K. Berndt, B. H. Hall, R. E. Hall, and J. A. Hausman. Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement,* 3(4):653–665, 1974.

[26] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics).* Springer-Verlag, Berlin, Heidelberg, 2006.

[27] J.-C. Breton. Processus stochastique. cours de M2 Mathématiques, Université de Rennes 1.

[28] K. P. Burnham and D. R. Anderson. *Model Selection and Inference: A Practical Information-theoretic Approach.* Intelligence, SS. of Lncs; 1501. Springer, 1998.

[29] B. Chapron. personal communication, 2018.

[30] K .L. Chung and J. B. Walsh. To Reverse a Markov Process. *Acta Mathematica,* 123:225–251, 1969.

[31] F. Comets and T. Meyre. *Calcul stochastique et modèles de diffusions.* Dunod, 2015.

[32] D. Commenges. Information theory and statistics: an overview.

[33] D. Commenges and H. Jacqmin-Gadda. *Dynamical Biostatistical Models.* Chapman & Hall/CRC Biostatistics Series. CRC Press, 2015.

[34] D. J. Crisp. The state-of-the-Art in Ship Detection in Synthetic Aperture Radar Imagery, 05 2004. Research Report.

[35] I. G. Cumming and F. H. Wong. *Digital Processing of Synthetic Aperture Radar Data.* Artech House, 2005.

[36] C. Dargatz. *Bayesian Inference for Diffusion Processes with Applications in Life Sciences.* PhD thesis, Ludwig Maximilian University of Munich, 2010.

[37] H. J. de Wind, J. E. Cilliers, and P. L. Herselman. DataWare: Sea Clutter and Small Boat Radar Reflectivity Databases. *IEEE Signal Processing Magazine*, 27(2):145–148, March 2010.

[38] *NIST Digital Library of Mathematical Functions.* http://dlmf.nist.gov/, Release 1.0.15 of 2017-06-01. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller and B. V. Saunders, eds.

[39] G. B. Durham and A. R. Gallant. Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes. *Journal of Business & Economic Statistics*, 20(3):297–316, 07 2002.

[40] T. Elfouhaily, B. Chapron, K. Katsaros, and D. Vandemark. A unified directional spectrum for long and short wind-driven waves. *Journal of Geophysical Research: Oceans*, 102(C7):15781–15796, 1997.

[41] T. Elfouhaily and C-A. Guérin. A critical survey of approximate scattering wave theories from random rough surfaces. *Waves in Random Media*, 14(4):R1–R40, 03 2004.

[42] J. H. G. Ender. Introduction to radar part 1. Lecture Notes, Ruhr-Universität Bochum.

[43] A. Farina, G. Gini, M. V. Greco, and L. Verrazzani. High resolution sea clutter data: statistical analysis of recorded live data. *IEE Proceedings - Radar, Sonar and Navigation*, 1997.

[44] P. Fayard and T. R. Field. Optimal inference of the scattering cross-section through the phase decoherence. *Waves in Random and Complex Media*, 18(4):571–584, 2008.

[45] P. Fayard and T. R. Field. Optimal inference of the inverse gamma texture for a compound-gaussian clutter. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2969–2972, April 2009.

[46] W. Feller. Two Singular Diffusion Problems. *Annals of Mathematics*, 54(1):173–182, 07 1951.

[47] T. Feng, T. R. Field, and S. Haykin. Stochastic Differential Equation Theory Applied to Wireless Channels. *IEEE Transactions on Communications*, 55(8):1478–1483, Aug 2007.

[48] T. R. Field. *Electromagnetic Scattering from Random Media*. Oxford University Press, 2009.

[49] T. R. Field and S. Haykin. Nonlinear Dynamics of Sea Clutter. *International Journal of Navigation and Observation*, 2008. Special Issue on Modelling and Processing of Radar Signals for Earth Observation.

[50] T. R. Field and R. J. A. Tough. Diffusion processes in electromagnetic scattering generating K-distributed noise. *The Royal Society*, 459:2169–2193, 2003.

[51] T. R. Field and R. J. A. Tough. Stochastic dynamics of the scattering amplitude generating K-distributed noise. *Journal of Mathematical Physics*, 44(11):5212–5223, 2003.

[52] D. Foata, J. Franchi, and A. Fuchs. *Calcul des probabilités*. Dunod, 2012.

[53] T. Funaki. *Lectures on Random Interface*. Springer, 2016.

[54] J.-F. Le Gall. *Mouvement brownien, martingales et calcul stochastique*. Springer, 2013.

[55] L. Gallardo. *Mouvement brownien et calcul d'Itô*. Editions Hermann, 2008.

[56] J. L. Garcia-Palacios. Introduction to the theory of stochastic processes and brownian motion problems. Lecture notes for a graduate course, Universidad de Zaragoza, 2004.

[57] F. Gini and M. Greco. Texture Modeling and Validation Using Recorderd High Resolution Sea Clutter data. *Proceedings of the 2001 IEEE Radar Conference*, 2001.

[58] A. Giroux. *Initiation à la mesure et à l'intégration*. Editions Ellipses, 2015.

[59] Y. E. Gliklikh. *Global and Stochastic Analysis with Applications to Mathematical Physics*. Springer, 2011.

[60] D. J. Griffiths. *Introduction to Electrodynamics*. Prentice Hall, 1999.

[61] U. G. Haussmann and E. Pardoux. Time Reversal of Diffusions. *The Annals of Probability*, 14(4):1188–1205, 1986.

[62] D. J. Higham. An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations. *Society for Industrial and Applied Mathematics*, 43(3):525–546, 2001.

[63] S. Iyanaga and Y. Kawasa. *Encyclopedic dictionary of mathematics*. MIT press, 1980.

[64] J. Jacod. On continuous conditional gaussian martingales and stable convergence in law. *Séminaire de Probabilités XXXI*, pages 232–246, 1997.

[65] J. Jacod and P. Protter. Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.*, 26(1):267–307, 01 1998.

[66] B. Jähne and K. S. Riemer. Two-dimensional wave number spectra of small-scale water surface waves. *Journal of Geophysical Research: Oceans*, 95(C7):11531–11546, 1990.

[67] E. Jakeman. On the statistics of K-distributed noise. *Journal of Physics A: Mathematical and General*, 13(1):31–48, 1980.

[68] E. Jakeman, K. I. Hopcraft, and J. O. Matthews. Distinguishing population processes by external monitoring. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 459(2031):623–639, 2003.

[69] E. Jakeman and P. N. Pusey. Significance of K Distributions in Scattering Experiments. *Phys. Rev. Lett.*, 40:546–550, Feb 1978.

[70] E. Jakeman and K. D. Ridley. *Modeling Fluctuations in Scattered Waves*. CRC Press: Series in Optics and Optoelectronics, 2006.

[71] E. Jakeman and R. J. A. Tough. Generalized K distribution: a statistical model for weak scattering. *Journal of the Optical Society of America A*, 4(9), September 1987.

[72] E. Jakeman and R. J. A. Tough. Non-gaussian models for the statistics of scattered waves. *Advances in Physics*, 37(5):471–529, 1988.

[73] E. Jay, J. P. Ovarlez, D. Declercq, and P. Duvaut. BORD: Bayesian Optimum Radar Detector. *Elsevier Signal Processing*, 83(6):1151–1162, 06 2003.

[74] A. Jayaprakash, G. Ramachandra Reddy, and N.S.S.R.K. Prasad. Small Target Detection Within Sea Clutter Based on Fractal Analysis. *Procedia Technology*, 24:988 – 995, 2016. International Conference on Emerging Trends in Engineering, Science and Technology (ICETEST - 2015).

[75] E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

[76] M. Jeong and J. Y. Park. Asymptotic theory of maximum likelihood estimator for diffusion model. Working paper, 2010.

[77] Samuel W. McCandless Jr. and Christopher R. Jackson. *Synthetic Aperture Radar Marine User's Manual, chapter 1*. http://www.sarusersmanual.com/.

[78] M. B. Kanevsky. *Radar Imaging of the Ocean Waves*. Elsevier, 2008.

[79] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. 2nd edition, Springer-Verlag, 1991.

[80] I. Karoui, I. Quidu, and M. Legris. Automatic Sea-Surface Obstacle Detection and Tracking in Forward-Looking Sonar Image Sequences. *IEEE Transactions on Geoscience and Remote Sensing*, 53(8):4661 – 4669, 08 2015.

[81] M. G. Kendall, A. Stuart, and J. K. Ord, editors. *Kendall's Advanced Theory of Statistics*. Oxford University Press, Inc., New York, NY, USA, 1987.

[82] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, 1992.

[83] G. J. Komen, L. Cavaleri, M. Donelan, K. Hasselmann, S. Hasselmann, and P. A. E. M. Janssen. *Dynamics and Modelling of Ocean Waves*. Cambridge University Press, 1994.

[84] P. Kumar and S. Narayanan. Solution of Fokker-Planck equation by finite element and finite difference methods for nonlinear systems. *Sadhana*, 31(4):445–461, Aug 2006.

[85] C. Lacaux. Probabilités. cours de probabilités, deuxième année option Ingénierie Mathématique, Ecole des Mines de Nancy.

[86] J.-C. Laleuf. *Processus et intégrales stochastiques*. Editions Ellipses, 2014.

[87] K. L. Lange, R. J. A. Little, and J. M. G. Taylor. Robust Statistical Modeling Using the t Distribution. *Journal of the American Statistical Association*, 84(408):881–896, 1989.

[88] S. Lessard. *Processus stochastiques*. Editions Ellipses, 2014.

[89] R. S. Lipster and A. N. Shiryaev. *Statistics of Random Processes, Volume 1, second edition*. Springer-Verlag, 2001.

[90] T. Lo, H. Leung, J. Litva, and S. Haykin. Fractal characterisation of sea-scattered signals and detection of sea-surface targets. *IEE Proceedings F - Radar and Signal Processing*, 140(4):243–250, 1993.

[91] F. Luo, D. Zhang, and B. Zhang. The Fractal Properties of Sea Clutter and Their Applications in Maritime Target Detection. *IEEE GEOSCIENCE AND REMOTE SENSING LETTERS*, 10(6):1295–1299, 2013.

[92] B. R. Mahafza. *Radar Systems Analysis and Design Using Matlab*. Chapman & Hall/CRC, 2000.

[93] A. T. Mai, F. Bastin, and M. Toulouse. On Optimization Algorithms for Maximum Likelihood Estimation. Technical report, CIRRELT, 12 2014.

[94] H. Maitre. *Processing of Synthetic Aperture Radar (SAR) Images*. Wiley-IEEE Press, 1 edition, 2008.

[95] M. E. Mancino and M. C. Recchioni. Fourier Spot Volatility Estimator: Asymptotic Normality and Efficiency with Liquid and Illiquid High-Frequency Data. *PLOS ONE*, 10(9):1–33, 09 2015.

[96] M. Martorella, F. Berizzi, and E. Dalle Mese. On the Fractal Dimension of Sea Surface Backscattered Signal at Low Grazing Angle. *IEEE Transactions on Antennas and Propagation*, 52(5):1193–1204, 2004.

[97] D. Massonet and J-C. Souyris. *Imaging with Synthetic Aperture Radar*. EPFL Press, 2008.

[98] J. H. Matis and T. R. Kiffe. *Stochastic Population Models: A Compartmental Perspective*. Springer, 2000.

[99] M. I. Mishchenko, L. D. Travis, and A. A. Lacis. *Multiple Scattering of Light by Particles: Radiative Transfer and Coherent Backscattering*. Cambridge University Press, 2006.

[100] C. D. Mobley. Modeling Sea Surfaces: A Tutorial on Fourier Transform Techniques. Technical report, Sequoia Scientific, Inc., 2016.

[101] P. Mörters and Y. Peres. *Brownian Motion*. Cambridge Series in Statistical and Probabilistic Mathematics, 2010.

[102] R. I. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley Series in Probability and Mathematical Statistics, 2005.

[103] J. N. Nielsen, H. Madsen, and P. C. Young. Parameter estimation in stochastic differential equations: an overview. *Annual Reviews in Control*, 2000.

[104] K. B. Nowman. Gaussian Estimation of Single-Factor Continuous Time Models of The Term Structure of Interest Rates. *The Journal of Finance*, 52(4):1695–1706, 09 1997.

[105] B. Oksendal. *Stochastic Differential Equations: An Introduction with Applications, Fifth Edition*. Springer-Verlag, 2000.

[106] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor. Complex Elliptically Symmetric Distributions: Survey, New Results and Applications. *IEEE Transactions on Signal Processing*, 60(11):5597–5625, 2012.

[107] S. Panagopoulos and J. J. Soraghan. Small-Target Detection in Sea Clutter. *IEEE Transactions on Geoscience and Remote Sensing*, 42(7):1355 – 1361, 2004.

[108] E. Pardoux. Stochastic Partial Differential Equations. Lectures given in Fudan University, Shanghaï, April 2007.

[109] T. Pham-Gia, N. Turkkan, and E. Marchand. Density of the ratio of two normal random variables and applications. *Communication in Statistics- Theory and Methods*, 35(9):1569–1591, 2006.

[110] P. C. B. Phillips and J. Yu. Maximum Likelihood and Gaussian Estimation of Continuous Time Models in Finance. *Handbook of Financial Time Series*, 2009.

[111] L. Pichler, A. Masud, and L. A. Bergman. *Numerical Solution of the Fokker–Planck Equation by Finite Difference and Finite Element Methods—A Comparative Study*, pages 69–85. Springer Netherlands, Dordrecht, 2013.

[112] W. J. Pierson and L. Moskowitz. A proposed spectral form for fully developed wind seas based on the similarity theory of S. A. Kitaigorodskii. *Journal of Geophysical Research*, 69(24):5181–5190, 1964.

[113] N. Pinel and C. Bourlier. *Electromagnetic Wave Scattering from Random Rough Surfaces: Asymptotic Models*. Wiley, 2013.

[114] E. Platen. An introduction to numerical methods for stochastic differential equations. *Acta Numerica*, 8:197–246, 1999.

[115] M. Podolskij and M. Vetter. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 2009.

[116] P. E. Protter. *Stochastic Integration and Differential Equations*. Springer, 2005.

[117] M. Rangaswamy. Spherically invariant random processes for modeling non-gaussian radar clutter. In *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, pages 1106–1110 vol.2, Nov 1993.

[118] N. J. Redding. Estimating the Parameters of the K distribution in the Intensity Domain. Technical report, ELECTRONICS RESEARCH LAB SALISBURY, Australia, 1999.

[119] M. Reiß. Statistics of stochastic processes. Lecture Notes a course at Humbold University of Berlin.

[120] H. Risken. *The Fokker-Planck Equation: methods of solution and applications.* Springer, 1989.

[121] C. J. Roussel, A. Coatanhay, and A. Baussard. Estimation of the parameters of stochastic differential equations for sea clutter. *IET Radar, Sonar and Navigation*, 13(4):497 − 504, 2018.

[122] C. J. Roussel, A. Coatanhay, and A. Baussard. Forward and backward probabilistic inference of the sea clutter. *Waves in Random and Complex Media*, 29(3):540–568, 2018.

[123] C. J. Roussel, A. Coatanhay, and A. Baussard. Detection of a coherent scatterer in a random medium: an approach based on transition probabilities. *submitted to Waves in Random and Complex Media*, 2019.

[124] T. Rylander, A. Bondeson, and P. Ingelström. *Computational Electromagnetics.* Texts in Applied Mathematics 51. Springer, second edition edition, 2013.

[125] H s. Choi. Information theory for maximum likelihood eestimation of diffusion models. *Journal of Econometrics*, 191(1):110–128, 2016.

[126] H s Choi, M. Jeong, and J. Y. Park. An asymptotic analysis of likelihood-based diffusion model selection using high frequency data. *Journal of Econometrics*, 178:539–557, 2014.

[127] R. Salmon. Introduction to Ocean Waves. Lecture Notes.

[128] A. D. Sanford and G. M. Martin. Bayesian comparison of several continuous time models of the australian short rate. *Accounting & Finance*, 46(2):309–326, 2006.

[129] J. Schwinger, L. L. DeRaad Jr., and K. A. Milton W. Tsai. *Classical Electrodynamics.* Advanced Book Program, 1998.

[130] M. T. Silva. Ocean surface wave spectrum. lecture notes.

[131] M. I. Skolnik. *Radar Handbook.* Electronic engineering series. McGraw-Hill, 1990.

[132] J. A. Stratton. *Electromagnetic Theory.* McGraw Hill, 1941.

[133] P. Tankov and E. Voltchkova. Jump-diffusion models: a practitioner's guide. *Banque et Marchés*, 99, 2009.

[134] F. Totir, E. Radoi, L. Anton, C. Ioana, A. Serbanescu, and S. Stankovic. Advanced sea clutter models and their usefulness for target detection. In *MTA REVIEW*, volume 18, 2008.

[135] R. J. A. Tough. A Fokker-Planck description of K-distributed noise. *Journal of Physics A: Mathematical and General*, 20(3):551–567, 1987.

[136] Leung Tsang, Jin Au Kong, Kung-Hau Ding, and Chi On Ao. *Scattering of Electromagnetic Waves: Numerical Simulations*. John Wiley & Sons, 2001.

[137] V. Vanaja. Numerical solution of a simple Fokker-Planck equation. *Applied Numerical Mathematics*, 9(6):533–540, 05 1992.

[138] H. C. von Baeyer. *QBism: the Future of Quantum Physics*. Harvard University Press, 2016.

[139] J. Wang, A. Dogandzic, and A. Nehorai. Maximum Likelihood Estimation of Compound-Gaussian Clutter and Target Parameters. *IEEE Transactions on Signal Processing*, 54(10):3884–3898, Oct 2006.

[140] K. D. Ward, C. J. Baker, and S. Watts. Maritime surveillance radar. I. radar scattering from the ocean surface. *IEE Proceedings F - Radar and Signal Processing*, 1990.

[141] K. D. Ward, R. J. A. Tough, and S. Watts. *Sea Clutter: Scattering, the K distribution and Radar Performance*. 20. IET Radar, Sonar and Navigation, first edition edition, 2006.

[142] K. D. Ward, R. J. A. Tough, and S. Watts. *Sea Clutter: Scattering, the K distribution and Radar Performance*. 25. IET Radar, Sonar and Navigation, second edition edition, 2013.

[143] G. N. Watson. *A Treatise on the Theory of Bessel Functions*. Cambridge University Press, 1995.

[144] S. Watts. Radar sea clutter: Recent progress and future challenges. In *2008 International Conference on Radar*, pages 10–16, Sep. 2008.

[145] K. Yao. *Spherically Invariant Random Processes: Theory and Applications*, pages 315–331. Springer US, Boston, MA, 2003.

[146] D. S. Zhang, G. W. Wei, D. J. Kouri, and D. K. Hoffman. Numerical method for the nonlinear Fokker-Planck equation. *Phys. Rev. E*, 56:1197–1206, Jul 1997.

[147] M. P. Zorzano, H. Mais, and L. Vazquez. Numerical solution for Fokker-Planck equations. In *Proceedings of the 1997 Particle Accelerator Conference (Cat. No.97CH36167)*, volume 2, pages 1825–1827 vol.2, May 1997.

[148] M. P. Zorzano, H. Mais, and L. Vazquez. Numerical solution of two dimensional Fokker-Planck equations. *Applied Mathematics and Computation*, 98(2):109 – 117, 1999.

Résumé : La télédétection radar en contexte maritime est généralement perturbée par les ondes radar réfléchies par la mer, contribution qui est appelée clutter de mer (ou fouillis) en raison de son caractère complexe voire aléatoire. Une compréhension approfondie du clutter de mer est nécessaire pour la détection de cible et l'imagerie radar. Des modèles statistiques ont longtemps été utilisés (K distribution, distribution de Weibull, etc), mais ils ont l'inconvénient d'être statiques. Nous proposons d'utiliser un modèle dynamique développé par T. R. Field, au sein duquel le clutter de mer est un processus aléatoire qui résout des équations différentielles stochastiques. Nous introduisons le modèle de Field pour la surface équivalente radar (SER) et le speckle. La réflectivité complexe de la mer dépend alors de trois paramètres : A, B et α.
Nous calculons les probabilités de transition de la SER et du speckle par résolution analytique d'équations de Fokker-Planck. Nous proposons alors de les utiliser comme outils pour synchroniser des observations prises à des positions et temps différents, comme dans le cas des Radars à Synthèse d'Ouverture.

Nous estimons les paramètres A et B par maximum de vraisemblance (MV) et montrons numériquement qu'il est possible à moindre coût d'approximer les probabilités de transition exactes par des gaussiennes grâce au schéma d'Euler-Maruyama. α est quant à lui estimé par ergodicité (moment). Nous adaptons le modèle de Field pour prendre en compte une cible simple, et montrons qu'il est possible d'estimer les paramètres de la cible par MV en utilisant à nouveau des approximations gaussiennes pour les probabilités de transition. Dans la dernière partie, nous abordons la non-observabilité de la SER en l'estimant à partir de la réfléctivité complexe (observable). Nous proposons un schéma composé d'une suite d'estimateurs directement applicable à des données réelles. Finalement, nous introduisons et discutons l'estimation bayésienne des paramètres du clutter, et la détection de cible, comme potentiels futurs travaux de recherche.

Radar remote sensing in a maritime context is often hindered by radar waves reflected by the sea, termed sea clutter due to its noise-like character. A thorough understanding of it is required for detection and imaging applications. Statistical models have long been used for the sea clutter (K distribution, Weibull distribution etc) but they are static in nature. We propose to use a dynamic model developed by T. R. Field, which represents the sea clutter as a stochastic process solving stochastic differential equations. We introduce Field's model for the sea surface radar cross section (RCS) and speckle. The complex reflectivity of the sea surface then depends on three parameters: A, B and α.
We compute the transition probabilities of the RCS and speckle by analytical resolution of Fokker-Planck equations, and propose to use them as a tool for synchronizing observations taken at differents positions and times, as in Synthetic Aperture Radar.

We derive maximum likelihood (ML) estimators for A and B, and show numerically that the exact transition probabilities from the Fokker-Planck equations can be approximated in a satisfactory manner by Gaussians using Euler-Maruyama's scheme. α, for its part, is estimated by ergodicity (moment). We adapt Field's model to account for the presence of a simple target and show that it is possible to estimate the target constant by ML using Gaussian approximations for the transition probabilities. In the last part, we address the non-observability of the RCS by estimating it from the complex reflectivity (observable). We obtain a sequence of estimators applicable to real data. Finally, bayesian estimation of the clutter parameters, and target detection, are introduced and discussed as potential future directions for research.