Forecasts of dynamical systems from analogs : applications to geophysical variables with a focus on ocean waves

Paul Platzer

Par
Paul PLATZER

Forecasts of dynamical systems from analogs

Applications to geophysical variables with a focus on ocean waves

Rapporteurs avant soutenance :

Julien Touboul   Maître de conférences, Institut Méditerranéen d'Océanographie - OPLC, Marseille
Dimitris Giannakis   Professeur associé, Courant Institute of Mathematical Sciences - CAOS, New-York

Composition du Jury :

Présidente :   Valérie Monbet   Professeure, Université Rennes 1
Examinateur·ice·s :   Anne-Catherine Favre   Responsable de l'équipe HMCI, INP Grenoble
                      Julien Touboul   Maître de conférences, IMO - OPLC, Marseille
                      Dimitris Giannakis   Professeur associé, CIMS - CAOS, New-York
Dir. de thèse :   Thierry Chonavel   Professeur, IMT Atlantique, Brest
Co-dir. de thèse :   Pierre Tandeo   Maître de conférences, IMT Atlantique, Brest
                      Jean-François Filipot   Directeur scientifique, France Énergies Marines, Brest
                      Pascal Yiou   Directeur de recherche, LSCE, Gif-sur-Yvettes
                      Philippe Naveau   Directeur de recherche, LSCE, Gif-sur-Yvettes

Invité :
Bertrand Chapron   Directeur de recherche, Ifremer, Brest

REMERCIEMENTS

La thèse est une chance. C'est un apprentissage unique, fascinant, très personnel et intense. C'est également un chemin long, parfois éprouvant, qu'on ne peut pas faire seul.

Tout d'abord, il faut être guidé. Bien qu'on ne sache jamais exactement où on va, il reste nécessaire de garder un certain cap, et d'être sûr qu'on arrivera quelque part. J'ai eu la chance d'être encadré par Jean-François, Pascal, Philippe, Pierre, et Thierry. Leur savoir et savoir-faire, leur perspicacité et leur modestie m'ont beaucoup appris, et continuent de m'impressionner et de m'inspirer. Leur soutien, leur attention et leur humanité m'ont permis d'aller au bout. Je ne peux pas imaginer de meilleures conditions de travail que celles qu'ils m'ont offertes. Merci.

Une fois le chemin parcouru, il faut atterrir. Anne-Catherine Favre, Dimitris Giannakis, Julien Touboul et Valérie Monbet m'ont fait le grand honneur de juger mes travaux écrits, ainsi que ma présentation orale de soutenance de thèse. Je n'oublierai jamais leurs remarques, questions, critiques et compliments. Je m'estime extrêmement privilégié d'avoir ainsi bénéficié de leur expertise et de leur attention. Les derniers mois de mon doctorat se sont extrêmement bien déroulés grâce à eux. Merci.

De nombreux autres professeurs et chercheurs m'ont également soutenu, épaulé et stimulé tout au long de ces trois ans: Pierre Ailliot, Bertrand Chapron, Marc Prevosto, Nicolas Raillard, Fabien Leckler, Rui Duarte, Benoît Saussol, Maxime Beauchamp, Ronan Fablet, Valérie Burdin. Merci.

Avant de commencer une thèse, il faut déjà le vouloir. Au cours de mes études, j'ai rencontré plusieurs professeurs et chercheurs qui m'ont donné l'envie de faire de la recherche par leur dévotion, leur érudition et leur humanité: Jean-Paul Goualard, François Métivier, Gilles Deruelle, Christophe Gissinger, Frédéric Chevy, Guilhem Semerjian, David Quéré, Pierre-Yves Lagrée, Jean-Martial Cohard, Guillaume Lapeyre, Marco Paukert, Corinna Hoose, Weeded Ghedaifi, et tant d'autres. Merci.

Tout au long de ces trois ans, j'ai été stimulé et aidé par de nombreux étudiants, collègues, amis, avec qui j'ai pu discuter de mes travaux parfois, mais surtout de tout et de rien: Alex, Alexis, Bastien, Maxime, Caio, MarieS, Audrey, Fred, Erwan, Duong, Saïd, Yicun, Sabrina, Noura, Simon, Valentin, Redouane, Thomas, NicolaS, AurélienS,

amour inconditionnel, aussi immuable que la Nature elle-même, un amour tel que celui que l'on ressent pour ses proches, sa famille. C'est ma manière de vous dire à toutes et tous, merci.

Je remercie encore toutes les personnes qui ont contribué aux vidéos de pré- et post-soutenance: Suzanne, Gérard, et tous ceux que j'ai déjà cité. C'était magique.

Enfin, je n'aurais rien réussi sans celle qui a été à mes côtés du début à la fin. Celle qui a partagé mes peines, calmé mes angoisses, qui a su trouver les mots lorsque je voulais arrêter. Celle qui, lorsque je donnais toute mon énergie et mon temps à mon travail, prenait son temps et son énergie pour faire ce que je ne faisais pas, et bien plus encore. Celle qui m'a donné l'envie d'avoir envie. Les vers de Stevie Wonder lui sont dédiés, comme aux autres membres de ma famille. Cette thèse, c'est un peu comme si on l'avait écrite à deux. Et puis, en vrai, t'es la meilleure découverte que j'ai faite en trois ans de recherches. Merci.

J'oublie forcément du monde... J'en suis sincèrement désolé. Sachez que je me rattraperais en payant des coups quand on aura de nouveau le droit d'en prendre.

À mes grands-parents,
Denise et Édouard Aspeele.

TABLE OF CONTENTS

RÉSUMÉ EN FRANÇAIS

Introduction (français)

Par un matin ensoleillé d'avril, une vieille dame et son petit-fils plantent des graines de courges dans le jardin. Lorsqu'il se rend compte qu'ils n'ont pas apporté d'arrosoir, le petit-fils commence à retourner vers la maison.
"Pas la peine" lui dit sa grand-mère, "il va pleuvoir cet après-midi de toute façon."
"Mais... ça fait trois jours qu'on a pas vu un nuage !"
"Il va pleuvoir, il pleut toujours quand le vent souffle dans cette direction."

Les personnes qui ont vécu longtemps à un endroit donné prétendent souvent qu'elles peuvent prédire la météo en se basant sur leurs observations passées. La méthode de prévision par "analogues" est utilisée en météorologie depuis les années 1970, et repose sur un principe similaire. Dans notre exemple, la grand-mère a déjà observé de nombreuses situations météorologiques "analogues" à celle du matin, et pour toutes ces situations passées le temps de l'après-midi (appelé "successeur") était pluvieux. Par conséquent, la grand-mère affirme que la probabilité d'une après-midi pluvieuse est élevée.

Les méthodes par analogues appartiennent à la catégorie des prévisions *par les données*, contrairement aux prévisions *par les modèles* qui se basent sur une description physique précise du système. Dans certains cas, une telle description physique est indisponible, ou bien elle impliquerait des calculs numériques trop lourds (et donc trop lents) pour des fins de prévision. Dans notre exemple, la grand-mère n'a pas eu besoin de décrire avec précision les relations entre la pluie et le vent; elle n'a pas non plus eu besoin de résoudre des équations compliquées à l'aide d'un ordinateur. Il lui a simplement suffi de se souvenir de ses observations passées. Dans le Chapitre I, je donne une revue de certains des avantages et inconvénients consubstantiels à l'utilisation de méthodes par analogues.

Durant cette thèse, j'ai tenté de répondre à la question suivante:

11

"Quand, et comment est-ce que les prévisions par analogues permettent une estimation précise de l'évolution d'un système géophysique ?"

Pour répondre à cette question, j'ai utilisé des modèles mathématiques pour rendre compte de certaines hypothèses sous-jacentes des méthodes par analogues.

Hypothèse n°1: "Il est possible de trouver de bons analogues."

Les méthodes par les données reposent sur la disponibilité d'une grande quantité de données, mais *combien de données* sont effectivement nécessaires pour les méthodes par analogues ? Dans notre exemple, la grand-mère semble avoir vécu assez longtemps pour trouver de nombreux analogues de la situation météorologique du matin, mais ce n'est pas le cas de son petit-fils. La question "combien de temps devons-nous attendre?" pour trouver de bons analogues a été étudiée parVan Den Dool (1994) et Nicolis (1998), montrant que les systèmes de grande complexité requièrent une grande base de données. Pour être plus spécifique, la quantté de données nécessaire dépende de manière exponentielle de la dimension du système. Par exemple, il est plus difficile de trouver des analogues de la situation météorologique sur l'entièreté du globe terrestre que sur une zone plus restreinte telle que la ville de Brest. Dans le chapitre II, j'étudie la question "ayant attendu un temps donné, quelle qualité d'analogues peut-on espérer?", qui est légèrement différente de la question des études précédentes: "combien de temps doit-on attendre pour trouver *un bon analogue*?". J'ai fais le choix d'une question reliée plus directement à l'usage pratique des analogues. En effet, il n'est en général pas possible "d'attendre plus longtemps", et une quantité finie de donnée est disponible. Par ailleurs, il est fréquent que l'application nécessite non pas un, mais bien plusieurs analogues.

Hypothèse n°2: "Le futur peut être estimé à partir des successeurs de bons analogues."

Cette hypothèse est reliée à l'idée que si deux états d'un système sont initialement proches, ils le resteront pendant un certain temps. Dans le chapitre III, j'utilise une description mathématique de la divergence entre les successeurs et le véritable état futur, en supposant que les analogues sont proches de l'état présent. Cette description permet de comparer les propriétés de méthodes existantes de prévision par analogues. Cet étude permet de choisir quelle méthode de prévision par analogue utiliser pour un problème donné. Elle garantit également le succès des prévisions par analogues dans le cadre général considéré, et en supposant que l'on a "attendu assez longtemps". D'autres études sur les garanties pour les prévisions par analogues ont été effectuées par Farmer and Sidorowich

(1987), Zhao and Giannakis (2016) et Alexander and Giannakis (2020).

La prévision par analogue n'estime pas uniquement l'état futur, mais aussi les incertitudes associées à cette prévision. Dans notre exemple, la grand-mère est confiante sur le fait qu'il va pleuvoir cet après-midi car elle a observé un grand nombre de situations analogues, toutes suivies par de la pluie. À l'inverse, si l'on a observé de nombreuses situations météorologiques analogues à la situation actuelle, et que les successeurs associés sont tous différents, la confiance associée à la prévision sera faible. Dans le chapitre III, je relie la variabilité dans le jeu de successeurs aux propriétés intrinsèques du système. Cela montre que l'estimation de l'incertitude de prévision par analogues est liée à la prévisibilité intrinsèque du système. Ces résultats sont en accord avec ceux de Atencia and Zawadzki (2017).

Hypothèse n°3: "Les hypothèses n°1 et n°2 sont toujours valables pour les évènements extrêmes."

La prévision d'évènements extrêmes revêt une importance particulière en ce qu'elle permet la mise en place de mesures de sécurité, permettant d'atténuer des effets potentiellement importants. De tels évènements sont rares, rendant la prévision par analogues délicate. Prenons l'exemple d'une vague océanique extrême, telle que décrite par Ernest Shackleton (Müller et al., 2005) lors de la fameuse expédition de 1 300km à bord du *James Caird*: *"À minuit, j'étais à la barre lorsque j'ai aperçu une fraction de ciel clair au sud, sud-ouest. J'ai prévenu l'équipage que le ciel se dégageait, puis soudain, j'ai réalisé que ce que j'avais vu n'était pas une brèche dans les nuages mais la crête blanche d'une énorme vague. En vingt-six ans de navigation parmi les océans dans tous leurs états, je n'avais pas rencontré de vague aussi gigantesque."*

Il est évidemment plus difficile de trouver des analogues d'un évènement qui n'est observé qu'une fois en 26 ans que d'évènements moins extrêmes. Ce raisonnement s'étend à d'autres méthodes basées sur les données. Par conséquent, il pourrait être intéressant d'utiliser des connaissances physiques pour prévoir de tels évènements. De nombreuses stratégies physiques ont été proposées pour la prévision de vagues océaniques extrêmes appelées "vagues scélérates" (Slunyaev, 2017). Dans le chapitre IV, je propose d'utiliser uniquement des mesures de vitesses de crêtes, plutôt que l'entièreté de la surface océanique. Une telle simplification permettrait de simplifier le processus de prévision.

Cependant, il reste important de comprendre les limites de l'utilisation des analogues pour la prévision des événements extrêmes. À quel moment les prévisions par analogues vont-elles échouer ? Dans le chapitre V, je montre des expériences numériques de prévi-

sions par analogues de variables aléatoires à queue lourde. Ces variables aléatoires sont utilisées pour modéliser certaines variables géophysiques telles que les précipitations horaires, qui ont une forte probabilité de s'écarter de leur valeur typique. Dans ces expériences, je montre que la taille des données nécessaire pour prévoir des événements extrêmes à l'aide d'analogues est plus importante que pour d'autres événements ordinaires. Ce constat justifie une adaptation des résultats théoriques du chapitre II au cas des événements extrêmes.

Résumé du chapitre 1

Le premier chapitre introduit les notions mathématiques et physiques nécessaires à la bonne compréhension de ce manuscrit. Elle est destinée aux lecteurs peu familiers avec la théorie des systèmes dynamiques, les analogues, les problèmes de filtrage, la théorie des valeurs extrêmes et les ondes de gravité à la surface de l'océan. Les références des livres sont indiquées pour les lecteurs qui souhaitent obtenir plus de détails. Pour les systèmes dynamiques, nous recommandons Katok and Hasselblatt (1997) et Cencini et al. (2010). Concernant le filtre particulaire, le livre de Cappé et al. (2006) est une référence, et nous conseillons Coles et al. (2001) pour une introduction à la théorie des valeurs extrêmes. De manière générale, la théorie des vagues gravitaires à la surface de l'océan peut être abordée à travers l'ouvrage de Kinsman (1965). Comme il n'existe pas, à notre connaissance, de livre de ce type pour les analogues, nous fournissons une introduction plus approfondie au sujet, ainsi qu'une revue de la littérature (non exhaustive). Nous en reprenons ici les éléments principaux.

Introduction aux analogues

On parle d'"analogue" lorsqu'un état d'un système est similaire à un autre état du même système. Cette similarité peut se caractériser par une faible distance au sens mathématique, ou par d'autres types de mesures de similarité. Dans la prévision par analogues, nous utilisons également les successeurs temporels des analogues. Les analogues sont donc des voisins dans l'espace des états, ou espace des phases, tandis que les "successeurs" sont leurs voisins dans le temps. La base de donnée dans laquelle les analogues sont cherchés est appelée "catalogue".

Le choix de l'espace dans lequel sont cherchés les analogues, ou "espace objet" (*feature*

14

space en anglais) est primordial. Il est le résultat d'un compromis entre la représentativité de l'espace, et la complexité de cet espace. Comme nous le verrons dans le chapitre 2, une trop forte complexité diminue les chances de trouver de bons analogues, tandis qu'une trop faible complexité empêche de rendre compte des phénomènes en jeu. L'usage des plongements temporels ("time-embeddings") est fréquent et permet de représenter l'état d'un système dans le cas d'observations partielles. Voir à ce sujet les études de Hamilton et al. (2016), Alexander et al. (2017), et Zhen et al. (2020).

Le choix de la distance qui caractérise les analogues est également fondamental. La distance Euclidienne étant très populaire et bien maîtrisée sur le plan théorique, de nombreux algorithmes (tels que les algorithmes en "arbres") ont été écrits pour accélérer la recherche d'analogues ou sayvoisins définis par cette distance. Après avoir effectué la recherche d'analogues par distance Euclidienne, d'autres mesures peuvent être utilisées, telles que la corrélation (Yiou et al., 2013) ou le score de Teweles-Wobus Blanchet et al. (2018). La distance de Mahalanobis (1936) est dérivée de la distance Euclidienne et permet de normaliser les données par leur structure de covariance, au prix d'un coût de calcul plus élevé. Fraedrich and Rückert (1998) et Wetterhall et al. (2005) adaptent des ditances dérivées de la distance Euclidienne pour un but précis, en ajustant poids pour optimiser la prévision ou la régionalisation (*downscaling*). La distance de Wasserstein permet de détecter des changements dans la structure des attracteurs de systèmes dynamiques (Robin et al., 2017). Cependant, cette distance est coûteuse, et est conçue à l'origine pour des distributions de probabilités, ce qui peut compliquer son utilisation pour des champs vectoriels à valeurs réelles (Thorpe et al., 2017).

Applications des analogues

Les analogues ont de nombreuses applications. Les premières utilisent une composante temporelle. Les analogues de circulation atmosphérique ont été introduits par Lorenz (1969) pour étudier la prévisibilité atmosphérique. Plus récemment, les analogues ont souvent été utilisés dans des schémas statistiques tels que le générateur stochastique de temps de Yiou (2014). L'utilisation d'analogues pour générer des prévisions d'ensemble a été comparée à d'autres techniques de génération d'ensembles par Delle Monache et al. (2013) et Atencia and Zawadzki (2017). La prévision par analogues a été combinée avec l'assimilation de données par Hamilton et al. (2016) et Lguensat et al. (2017), en utilisant soit des ensembles soit des hypothèses gaussiennes pour estimer la distribution des prévisions par analogues. Grooms (2020) a combiné la prévision par analogues avec

l'assimilation de données et les auto-encodeurs pour compenser un éventuel manque de données (lorsque le catalogue est trop petit). Sévellec and Drijfhout (2018) effectuent une discrétisation des valeurs possibles de température moyenne globale de surface, permettant d'effectuer des prévisions rapides avec une matrice de transition et des analogues des trajectoires CMIP5. En combinant les analogues, les noyaux gaussiens et les plongements temporels ("time-embeddings"), Alexander et al. (2017) effectue des prévisions des oscillations intrasaisonnières tropicales, basées sur un cadre introduit par Zhao and Giannakis (2016). Plus récemment, Wang et al. (2020) a utilisé des analogues dans un cadre généralisé de théorie des opérateurs (détaillé par Alexander and Giannakis, 2020) pour prévoir le phénomène d'oscillation El Niño.

De nombreuses applications des analogues n'impliquent pas d'évolution dans le temps (c'est-à-dire qu'elles n'utilisent que des analogues et non leurs successeurs). La plus connue est la régionalisation (downscaling) par analogues, qui consiste à estimer, à partir d'informations à grande échelle, les états possibles d'une station météorologique (impliquant des processus locaux à petite échelle). Par exemple, Wetterhall et al. (2005) produit des statistiques sur les précipitations dans des stations en Suède à partir d'analogues des champs de pression au niveau de la mer à grande échelle et des valeurs de précipitations locales associées. Les analogues permettent également d'effectuer l'opération inverse : à partir d'informations locales, la probabilité de situations météorologiques à grande échelle est estimée statistiquement de manière similaire. Cela permet de reconstruire des champs à haute résolution à des dates pour lesquelles seules des observations locales erronées sont disponibles (Schenk and Zorita, 2012; Yiou et al., 2013). De même, les analogues peuvent être utilisés pour spécifier une variable donnée à partir des analogues d'une autre variable (Tandeo et al., 2016). Enfin, le rôle de la circulation atmosphérique dans l'apparition d'événements extrêmes peut être estimé à l'aide d'analogues. Lorsqu'on observe un événement tel qu'un hiver froid (Cattiaux et al., 2010), une augmentation de la fonte de la surface de la calotte glaciaire (Fettweis et al., 2013), ou une vague de chaleur (Jézéquel et al., 2018), on peut trouver des analogues des circulations atmosphériques respectives et comparer les températures associées, ou la fonte de la glace. Cela permet d'estimer la probabilité d'observer des événements extrêmes en supposant des schémas de circulation similaires.

Avantages et inconvénients

Un avantage évident de l'utilisation des analogues est qu'elle permet de contourner l'utilisation d'un modèle physique. En particulier, les modèles physiques pour l'évolution temporelle d'observables spécifiques (par exemple, l'indice ENSO) ne sont généralement pas fournis ou sont difficiles à écrire. Enfin, même lorsqu'un modèle physique est disponible, sa résolution numérique peut être coûteuse. En outre, dans de tels modèles, les processus non linéaires impliquent des interactions entre toutes les échelles physiques, ce qui nécessite la paramétrisation des processus non résolus à petite échelle (en raison de la discrétisation). Il en résulte des incertitudes qui peuvent être difficiles à évaluer.

Les méthodes par analogues ont l'avantage d'être *peu coûteuses en termes de calcul*, de sorte qu'il est *facile d'effectuer des prévisions d'ensemble*, ce qui permet d'évaluer les incertitudes. Les librairies de calcul dédiées à l'apprentissage automatique permettent une recherche rapide des analogues, avec des arbres calculés hors ligne (voir l'annexe de Lguensat et al., 2017), et l'application des méthodes par analogues consiste généralement en de simples combinaisons linéaires. Au contraire, lorsque l'on utilise des méthodes basées sur des modèles issus de la connaissance physique, on ne peut espérer réaliser de très grandes prévisions d'ensemble en raison du coût élevé du calcul.

Les méthodes par analogues ont l'avantage d'être *non-paramétriques*. Ce qui signifie que seul un petit nombre de méta-paramètres doit être défini à l'avance (le nombre d'analogues utilisés à chaque itération, le choix de la distance, ...etc.) D'autres techniques doivent estimer les paramètres, en assumant une certaine forme ou un certain modèle fonctionnel. En particulier, lorsqu'on utilise des analogues dans un cadre statistique, il n'y a pas d'hypothèse sur la forme des distributions de probabilité. Il convient de noter que les méthodes par analogues peuvent également être utilisées sous une forme paramétrique lorsque cela est pertinent : voir la combinaison des analogues et du filtre de Kalman de Lguensat et al. (2017) ou l'optimisation des poids susmentionnée de Fraedrich and Rückert (1998) et Wetterhall et al. (2005).

Un autre point fort des méthodes par analogues est qu'elles *préservent les caractéristiques du système* par construction. Atencia and Zawadzki (2017) a noté que, contrairement aux autres méthodes de perturbation des conditions initiales, les prévisions d'ensemble par analogues sont cohérentes avec la dimension fractale du système. Zhen et al. (2020) note que la technique d'assimilation des données par analogues permet de construire des cartes d'élévation de surface en mer avec un spectre temporel plus cohérent avec les spectres réels qu'une interpolation optimale. Le générateur stochastique

17

de temps de Yiou (2014) préserve les contraintes spatiales par construction, contrairement aux autres générateurs stochastiques de temps.

La simplicité des méthodes par analogues les rend facilement *interprétables*, tandis que d'autres méthodes basées sur des données peuvent être considérées comme des "boîtes noires" où il est difficile de comprendre pourquoi et quand la méthode réussit ou échoue. Dans le cas de la prévision par analogues, la *convergence* vers une prévision optimale dans la limite de grandes données est prouvée par Zhao and Giannakis (2016) pour la prévision par analogues avec des noyaux adaptés à la dynamique. Dans le troisième chapitre de cette thèse et dans Platzer et al. (2019), nous développons des preuves similaires pour la prévision par analogues dans l'espace des phases. Enfin, Alexander and Giannakis (2020) établit des résultats généraux pour la prévision par analogues combinée avec des méthodes de projection de noyaux, faisant appel à la théorie des opérateurs de Koopman.

L'une des faiblesses des méthodes par analogues est qu'elles sont soumises à la *malédiction de la dimension* : le nombre de données nécessaires pour que les méthodes soient efficaces croît de manière exponentielle avec la dimension. Cette question est étudiée en détail dans le chapitre II, et les conséquences pour les prévisions par analogues sont examinées dans le chapitre III. Cependant, la malédiction de la dimension est une préoccupation générique pour les méthodes basées sur les données. En outre, comme indiqué précédemment, il s'agit maintenant de choisir correctement l'espace objet (feature space) afin de contourner ce problème.

Résumé du chapitre 2

Le chapitre II porte sur les distances entre les analogues et les états cibles. Il donne une description quantitative de l'hypothèse selon laquelle il est possible de trouver des analogues qui sont "proches" d'un état cible donné. Par rapport aux travaux précédents sur le sujet, le point de vue est modifié, passant de statistiques de temps de retour à des statistiques de distances d'analogues.

Nous donnons de nouvelles distributions de probabilité qui sont applicables au-delà du premier ou "meilleur" analogue. Celles-ci permettent d'estimer la probabilité de trouver de bons analogues d'un état donné, pour un système dynamique donné. La méthode proposée est basée sur la dimension locale et la taille du catalogue dans lequel les analogues sont recherchés. J'ai montré comment ces distributions de probabilité peuvent être utiles pour tout type de méthode par analogues, y compris la prévision. En particulier, la variabilité

des distances analogiques par rapport à la cible semble être une fonction décroissante de la dimension. Par conséquent, dans les problèmes de dimensions élevées, les meilleurs analogues d'un point donné sont "tous aussi bons les uns que les autres", alors que dans les dimensions basses, le premier meilleur analogue est susceptible d'être beaucoup plus "meilleur" que le 30ème meilleur analogue, par exemple.

Ces résultats sont la conséquence directe de travaux récents combinant des systèmes dynamiques et la théorie des valeurs extrêmes. Des tests sur des simulations numériques du système tridimensionnel de Lorenz et sur des cartes de vent issues d'un modèle physique confirment l'applicabilité de nos distributions de probabilité analytiques à des données géophysiques réelles.

Résumé du chapitre 3

Le chapitre III soutient l'utilisation des analogues à des fins de prévision et compare différentes stratégies de prévision par analogues, s'appuyant sur des arguments théoriques et des simulations numériques. Il peut être considéré comme un effort d'interprétation de méthodes existantes de prévision par analogues. Il est en cours de révision pour être publié dans le *Journal of Atmospheric Sciences*, et comprend des éléments d'un article tiré des actes de la conférence *Climate Informatics* de 2019.

Dans ce chapitre, j'ai donné une interprétation des erreurs de prévision par analogues en utilisant le flot du système. Cette interprétation vise à établir un lien entre les prévisions par analogues, basées sur des données, et les prévisions basées sur des modèles. Nous avons démontré que les erreurs moyennes de prévision par analogues dépendent linéairement de la distance entre l'état analogue et l'état initial, et quadratiques pour les méthodes par analogues qui utilisent des régressions linéaires locales (c'est-à-dire qui utilisent un opérateur localement linéaire). Il a été démontré que ces régressions linéaires permettent en fait d'estimer la matrice jacobienne du flot du système réel. J'ai interprété les opérateurs "locally-constant" et "locally-incremental" (LI) comme des cas particuliers de l'opérateur "locally-linear", lorsque le flot est soit constant dans l'espace des phases, soit égal à la matrice identité. Cela explique la précision de l'opérateur LI dans des délais courts. Ces résultats analytiques ont été confirmés par des expériences numériques de systèmes dynamiques chaotiques connus. La robustesse de ces résultats au bruit additif a également été examinée théoriquement et numériquement, justifiant l'utilisation d'un plus grand nombre d'analogues pour atténuer l'influence du bruit. Enfin, l'avantage éventuel

de diviser une prévision en plusieurs prévisions par analogues a été examiné, montrant des résultats préliminaires prometteurs lorsque les analogues sont combinés avec une régression linéaire.

Résumé du chapitre 4

Le chapitre IV ouvre une nouvelle perspective pour la prévision des vagues océaniques extrêmes : la focalisation des groupes de vagues individuels est prédite en utilisant uniquement les vitesses de crête. Cette nouvelle méthodologie est basée sur des démonstrations analytiques pour des groupes de vagues unidirectionnels, gaussiens et linéaires. Elle est testée sur des simulations numériques de vagues linéaires et de l'équation de Schrödinger non linéaire. Ce travail a été publié dans *Natural Hazards*.

Dans ce chapitre, je me suis concentré sur les vagues océaniques de grande amplitude appelées "vagues scélérates" qui constituent une menace pour les navires et les structures offshore. Une méthode basée sur un modèle pour la prévision de telles vagues a été présentée. J'ai montré que les futures plus hautes vagues pouvaient être prévues en utilisant les vitesses de crête plutôt que des champs de vagues entiers. Cela donnerait un avantage en termes de temps de calcul, en se concentrant uniquement sur les plus grandes vagues. Cependant, l'estimation des vitesses de crête à partir des techniques récentes de mesure des vagues reste un défi. Nos démonstrations analytiques établissent des profils de vitesses de crête en supposant une surface océanique linéaires, des spectres de vagues à bande étroite et des groupes de vagues gaussiens. Ces résultats ont été confirmés par des expériences numériques de vagues linéaires, et une méthode de prévision de la position et de l'amplitude du paquet de vagues focalisé a été testée. Les limites dues aux interactions non-linéaires des vagues ont été examinées dans des simulations numériques de l'équation non linéaire de Schrödinger.

Résumé du chapitre 5

Le chapitre V présente des études préliminaires sur l'utilisation d'analogues et de filtres particulaires pour la prévision de variables aléatoires à queue lourde. Un modèle jouet unidimensionnel avec des variables aléatoires à forte queue a été construit, sur lequel des méthodes analogiques ont été testées en combinaison avec des techniques de filtrage particulaire. Pour les variables aléatoires avec un indice de queue modéré, les prévisions

par analogues atteingnent l'efficacité des prévisions basées sur un modèle. La taille du catalogue nécessaire pour trouver des analogues d'événements extrêmes est apparue plus importante que ce qui était attendu d'après les résultats du chapitre II. Cela nécessite une adaptation de la théorie du chapitre II pour les distances d'analogues aux événements extrêmes, en tenant compte des paramètres de queue de la distribution.

INTRODUCTION

On a sunny morning of April, an old lady and her grand son are planting squash seeds in the backyard. When he realizes that they haven't brought a watering can, the grand son starts going back to the house.
"Don't bother" says his grand mother, "it's going to rain this afternoon anyway."
"But... it's been all sunny for three days now!"
"It will rain, it always does when the wind is blowing from this direction."

People who have lived for a long time at a given place commonly claim that they can predict the weather based on what they have already observed. A forecasting method called "analogs" have been used in meteorology since the 1970s, and is based on a similar principle. In our example, the grand mother has seen numerous "analog" situation that resemble this morning's weather, and for all these past situations, the afternoon weather or "successor" was rainy. Therefore, she claims that the probability to have rain in the afternoon is high.

Analog methods belong to the general class of *data-based* forecasts, in opposition to *model-based* forecasts which rely on a thorough physical description of the system. In some situations, such a physical description is either not available, or involves numerical computations that are too slow for forecast purposes. In our example, the grand mother did not have to describe precisely the relationship between wind and rain, nor did she have to solve complicated equations with the aid of a computer. It was enough for her to simply recall what she had observed in the past. In Chapter I, I give a review of some of the advantages and drawbacks of using analog methods.

In this thesis, I focused on the following question:

"How and when do simple forecasts based on analogs accurately estimate the evolution of a geophysical system?"

To answer this question, I used mathematical models to describe some of the underlying assumptions of analog methods.

Assumption n°1: "Good analogs can be found."

Data-based methods rely on the availability a large amount of data, but *how large*? In the example above, the grand mother seems to have lived long enough to find numerous analogs of today's weather, but that was not the case of her grand son. The question of "how long must we wait?" when searching for analogs has been treated by Van Den Dool (1994) and Nicolis (1998), showing that systems of high complexity demand a very large database. To be more precise, the number of data that is needed to find one good analog depends exponentially on the dimension of the system. For instance, it is harder to find analog situations of the weather on the entire globe than analogs of the weather on a smaller area (say, the city of Brest). In Chapter II, I tackle the question "if we have waited a given time, how good will *the best analogs* be?", which is different from the question of previous studies "how long must we wait to find *one good analog*?". The question I chose is more closely related to the practical use of analogs. Indeed, one is generally provided with a given amount of data and cannot hope to "wait longer". Also, one generally needs more than just one analog.

Assumption n°2: "The future can be estimated from the successors of good analogs."

This assumption is related to the notion that if two states of a system are initially close, they will remain so for a given amount of time. In Chapter III, I use a mathematical description of how successors diverge from the real future state, assuming that the analogs are close to the present state, which allows to compare the properties of some existing analog forecasting strategies. This could help to choose which analog method is appropriate to a given problem. It also guarantees that analog forecasts will suceed in the cases considered here, and "if we have waited long enough". Other studies on guarantees for analogs-like forecast strategies include the ones of Farmer and Sidorowich (1987), Zhao and Giannakis (2016) and Alexander and Giannakis (2020).

Analog forecasting not only estimates the future state, but also the uncertainties associated with such a forecast. In our example, the grand mother is confident that rain will fall in the afternoon because she can remember a large number of analog situations, all of which were followed by rainfall. Conversely, if one remembers many analog situations of the present weather, but if the associated successors are all different, one will not be confident in the estimation of the future weather. In Chapter III, I relate the variability in the set of successors to properties of the system. It shows that the uncertainty estimate of analog forecasting is related to the intrinsic predictability of the system. These findings are in agreement with another study by Atencia and Zawadzki (2017).

Assumption n°3: "Assumptions n°1 and n°2 hold for extreme events."

The forecast of extreme events is of particular importance because it can allow to take safety measures, mitigating potentially large impacts. Such events are usually rare, such that analogs may not succeed to forecast them. Let us take the example of an extreme ocean wave, as recounted by Ernest Shackleton (Müller et al., 2005) during the epic 1 300 km-long journey of the *James Caird*:

"At midnight, I was at the tiller and suddenly noticed a line of clear sky between the south and south-west. I called to the other men that the sky was clearing, and then a moment later I realized that what I had seen was not a rift in the clouds but the white crest of an enormous wave. During twenty-six years' experience of the ocean in all its moods I had not encountered a wave so gigantic."

Searching for analogs of an event that is observed only once every twenty-six years might be more difficult than for other, less extreme events. This reasoning applies to other data-based methods as well. Therefore, it might be interesting to use physical knowledge for such events. Many strategies based on physical models have been proposed for the forecast of extreme ocean waves called "rogue waves" (Slunyaev, 2017). In Chapter IV, I propose to use only measurements of wave crest velocities, rather than the whole sea surface. This could be an advantage as it would simplify the forecasting process.

However, it is important to understand the limitations of using analogs for the forecast of extreme events. At which point are analog forecasts going to fail? In Chapter V, I give numerical experiments of analog forecasts of heavy-tailed random variables. These random variables are used to model some geophysical variables such as hourly rainfall, which have a high probability of deviating far from their typical value. In these experiments, I show that the expected size of data that is needed to forecast extreme events using analogs is larger than for other "regular" events. This calls for an adaptation of the theoretical results of Chapter II to the case of extreme events.

Chapter-by-chapter content description

Chapter I introduces mathematical and physical notions on which the thesis is based. It is intended for readers unfamiliar with dynamical systems theory, analogs, filtering problems, extreme value theory, and ocean surface gravity waves. Books references are indicated to readers seeking for more details. As there is, to our knowledge, no such

book for analogs, we provide a more thorough introduction to the topic, along with a (non-comprehensive) litterature review.

Chapters II, III, IV, and V show analyses in the form of journal articles, and can be treated as independent research pieces. For each of these chapters, the first section outlines the scientific questions that are adressed, and the methods that are used. The main findings are summarized in each chapter's last section.

Chapter II focuses on the distances between analogs and target states. It gives a quantitative description of the assumption that one can find analogs that are "close" to a given target state. Compared to previous works on the matter, the point of view is changed from return time statistics to analog distance statistics. We give new, full probability distributions that are not limited to the first or "best" analog, but applicable to a finite number of analogs. These distributions can be used to estimate the accuracy of analog methods, including analog forecasting. Our results are examined in numerical simulations and on environmental datasets.

Chapter III supports the use of analogs for forecasting purposes and compares different analog forecasting strategies based on theoretical arguments and numerical simulations. It can be seen as an interpretation effort of some existing analog methods. It is under revision for publication in the *Journal of Atmospheric Sciences*, and includes elements of an article from the proceedings of the 2019 *Climate Informatics* conference.

Chapter IV opens a new perspective for the forecasting of extreme ocean waves: the focusing of individual wave packets is predicted using only crest velocities. This new methodology is based on analytical derivations for unidirectional, Gaussian, linear wave packets. It is tested on numerical simulations of linear waves and of the non-linear Shcrödinger equation. This work has been published in *Natural Hazards.*

Chapter V presents preliminary investigations on the use of analogs and particle filters for the forecast of heavy-tailed random variables. Numerical experiments of one-dimensional state-space models with heavy-tailed variables show that the number of data necessary for analog methods to recover model-based forecasts must be increased beyond the predictions of Chapter II.

The conclusion chapter recalls the main results of the thesis and draws perspectives for further research.

MATHEMATICAL AND PHYSICAL FRAMEWORKS

Preamble

The work presented here is multidisciplinary. Notions from various mathematical areas such as dynamical systems, statistics, and extreme value theory are used to study different physical variables such as wind and ocean waves. To better understand each chapter of this thesis, a reminder of key notions is necessary. The reader may skip this introductory chapter if already familiar with these notions.

The first section is devoted to dynamical systems. We first introduce the concept of flow, which is central to the study of analog forecasts presented in Chapter III. Then, since the divergence of trajectories from chaotic dynamical systems is an important issue for their forecast, we give a succinct introduction to Lyapunov exponents (the later are invoked several times in Chapter III). We then briefly recall theoretical results on the recurrence of chaotic dynamical systems that motivate the search for analogs. Finally, the ergodic, probabilistic description of chaotic dynamical systems is presented. This framework allows us to study the probability to find good analogs in Chapter II.

The second section is a general introduction to analog methods. We outline the vocabulary of analogs, always referring to practical aspects of using of analog methods. We then give a quick literature review of the many geophysical applications of analogs. To place analog methods in a wider context of statistical methods, some of their strength and weaknesses are reminded.

The third and fourth sections may help the novice reader to understand Chapter V. In this last chapter, we use a data assimilation technique called "particle filtering", of which we recall the principles here. Also, as this last chapter is concerned with the forecast of extreme events, it is necessary to introduce the framework of extreme value theory.

Finally, the fifth section outlines important aspects of the physics of ocean waves. In

Chapter IV, we explore the possibility to forecast large ocean waves from measurements of crest velocities, assuming that these waves originate from the focusing of linear, unidirectional, Gaussian wave packets. The physical foundations of these assumptions are recalled here for the novice reader.

1 Dynamical systems

For a more detailed introduction to dynamical systems theory, the reader is referred to the books of Katok and Hasselblatt (1997) and Cencini et al. (2010).

Note that a rigorous description of dynamical systems theory is out of the scope of this thesis. Therefore, concepts will be introduced intuitively, simplest case scenarios will be assumed, and thorough descriptions of the underlying mathematical assumptions will be eluded.

1.1 Flow

In many geophysical applications, the time-evolution of the system can be described by an equation of the type:

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{f}(\mathbf{x})$$

where t is time, the *state* $\mathbf{x}$ is a vector of phase-space $\mathcal{P}$, and $\mathbf{f}$ is a map $\mathcal{P} \mapsto \mathcal{P}$ that defines the dynamical system. The *phase-space* is such that a point $\mathbf{x} \in \mathcal{P}$ fully characterizes the state of the system[1]. Although more general types of manifold could be examined, we study here the simple case of n-dimensional Euclidean spaces $\mathcal{P} = \mathbb{R}^n$. We assume that $\mathbf{f}$ is both deterministic and independent of time. When considered, stochasticity will be expressed separately.

Integrating this differential equation from time t to time $t + l$ defines the map $\mathbf{\Phi}^l$

$$\mathbf{\Phi}^l(\mathbf{x}_t) = \mathbf{x}_{t+l}$$

that depends both on l and $\mathbf{x}_t$. More generally, one defines the *flow* $\mathbf{\Phi} : \mathbb{R} \times \mathcal{P} \mapsto \mathcal{P}$; $(l, \mathbf{x}_t) \mapsto \mathbf{x}_{t+l}$ of the dynamical system. The latter can be studied through either $\mathbf{f}$ or

---

1. For instance, in the case of an oscillating pendulum, a point of phase-space is a two-dimensional vector whose coordinates are the position and velocity of the pendulum. In physical space, Newton's law involves a second-order time-derivative, but the latter becomes first-order once expressed in phase-space.

$\mathbf{\Phi}$. Both approaches are used in Chapter III.

In geophysical applications, the state $\mathbf{x}_t$ is a high-dimensional[2] vector containing the values of many variables (e.g. temperature, pressure, wind, humidity) at every point of a discretized time-space grid. These variables obey physical laws that are expressed in the form of partial differential equations such as the Navier-Stokes equation of fluid flows, equations of heat transfer, phase change... These equations define the map $\mathbf{f}$, which further defines $\mathbf{\Phi}$ through time-integration.

1.2 Divergence of trajectories (chaos)

For chaotic dynamical systems, *initially close states diverge* exponentially in time under the action of the flow[3]. This sensitivity to initial conditions can be studied through *Lyapunov exponents*. A perturbation at time $t = 0$, noted $\delta\mathbf{x}_0$, grows exponentially at a rate $1/\lambda$ where λ is a Lyapunov exponent:

$$\|\delta\mathbf{x}_t\| \approx e^{\lambda t}\|\delta\mathbf{x}_0\|,$$

where $\delta\mathbf{x}_t = \mathbf{\Phi}^t(\mathbf{x}_0 + \delta\mathbf{x}_0) - \mathbf{\Phi}^t(\mathbf{x}_0)$. This scaling is valid for infinitesimally small $\|\delta\mathbf{x}_0\|$. The exponent λ depends on the position of $\mathbf{x}_0$, except for ergodic systems that will be introduced in the next section. λ also depends on the direction of $\delta\mathbf{x}_0$, so that there are as many exponents as the number of phase-space directions (a spectrum of Lyapunov exponents).

As long as there exists one positive Lyapunov exponent, the system is *unstable*, and any perturbation in the direction associated to that exponent will grow exponentially. In this case, the *Lyapunov time* is defined as the inverse of the largest Lyapunov exponent, and is the characteristic timescale at which perturbations in initial conditions grow. In practice the initial perturbation is finite, and the error $\|\delta\mathbf{x}_t\|$ grows exponentially for a time of the order of the Lyapunov time before reaching a plateau, especially for bounded nonlinear systems.

Chapter III is concerned with the forecast of analogs, and uses a linear approximation for the growth of the analog-to-future state distance, which involves the Jacobian matrix of the flow. There are strong connections between the Lyapunov exponents and the Jacobian matrix of the flow, although these connections are not expressed in detail here.

---

2. Infinite-dimensional in theory, but finite-dimensional approximations are used in practice.

3. Note that other definitions of chaos may be used, but we only report this simple notion of sensitivity to initial conditions.

More generally, the concept of Lyapunov time is central in the field of dynamical systems forecast, and will be invoked several times throughout Chapter III.

Systems that have a positive Lyapunov exponent are termed *chaotic*. A small misspecification of initial condition leads to a large forecast error, which is one of the reasons why systems like the atmosphere are hard to forecast. A famous quote on chaos attributed to Edward Lorenz is the following: "the present determines the future, but the approximate present does not approximately determine the future". This apparent randomness motivates the use of a *probabilistic description of trajectories*. In particular, their long-term behaviour can be described using ergodic theory, of which we now recall a few important results.

1.3   Recurrences and ergodicity

Poincaré's recurrence theorem states that, under the action of the flow, almost all points in any initial subset of the phase-space eventually revisit this subset (we will see later what "almost" means). In particular, the system will come back infinitesimally close to any initial point, and will do so an infinite number of times. This property of recurrence feeds the hope to find analogs of dynamical systems. The theorem was proven by Poincaré (1890) under two main hypotheses: the system can only access a finite volume of phase-space, and phase-space volumes are preserved under the action of the flow. These hypotheses are related to the notions of attractor and measure.

For many systems, including geophysical systems such as the atmosphere, long-term trajectories are organized to form a unique geometric object (which may be complex). Long-term trajectories[4] converge to a subset of phase-space $\mathcal{A} \subset \mathcal{P}$ called the *attractor*[5]. This has practical consequences for analog forecasting (see Chapter III). Otherwise stated, for any starting point $\mathbf{x}_0 \in \mathcal{P}$, there is a finite time τ after which all future states are in the attractor: $\{\mathbf{\Phi}^t(\mathbf{x}_0)\,,\ t > \tau\} \subset \mathcal{A}$.

However, some regions of the attractor may be visited more often than others. This motivates the definition of a measure μ on the subsets of $\mathcal{P}$ that characterizes how often each region of phase-space is visited by long-term trajectories of the system. Such a measure estimates the fraction of phase-space volume that is occupied by the attractor. It is more general then a probability density function. In the following, we assume that

---

4. To be a little bit more precise: "almost all" trajectories from the "basin of attraction" of the attractor converge to it.

5. Some systems may have more than one attractor. Here, we only deal with the simplest case.

the system has an attractor that can be characterized by such a measure. As this measure describes long-term trajectories, it must be invariant under the flow: for any $t > 0$ and $E \subset \mathcal{P}$, $\mu(\mathbf{\Phi}^t(E)) = \mu(E)$. Therefore, it is called the *invariant measure*. In ergodic theory, a property is said to be true for "almost all points" if the set of points that do not verify the property has zero measure.

This framework allows to state the *ergodic theorem* of Birkhoff (1931). This theorem shows that the statistics over all accessible states can be approximated by statistics over a very long trajectory, *whatever the starting point.* It has very important practical consequences in physics. Consider that we sample a trajectory of the system every Δt, from a starting point $\mathbf{x}_0 \in \mathcal{P} : \{\mathbf{x}_0, \mathbf{x}_{\Delta t}, \mathbf{x}_{2\Delta t}, \mathbf{x}_{3\Delta t}, \ldots\} = \{\mathbf{x}_0, \mathbf{\Phi}^{\Delta t}(\mathbf{x}_0), \mathbf{\Phi}^{2\Delta t}\mathbf{x}_0, \mathbf{\Phi}^{3\Delta t}\mathbf{x}_0, \ldots\}$. The ergodic theorem states that the average over this trajectory of any real-valued function of phase-space $g : \mathcal{P} \mapsto \mathbb{R}$ (also called a real-valued *observable*) tends to the average with respect to the invariant measure of the attractor:

$$\int_{\mathcal{P}} g \, \mathrm{d}\mu = \lim_{N \to \infty} \frac{1}{N} \sum_{k=0}^{N-1} g\left(\mathbf{\Phi}^{k\Delta t}(\mathbf{x}_0)\right) ,$$

for almost all starting points $\mathbf{x}_0$. A consequence of this theorem is that the measure of a given subset $E \subset \mathcal{P}$ can be approximated by the number of visits of this subset :

$$\mu(E) \approx \frac{1}{N} \sum_{k=0}^{N-1} \mathbb{1}\left(\mathbf{\Phi}^{k\Delta t}(\mathbf{x}_0) \in E\right) ,$$

for any starting point $\mathbf{x}_0$ and a sufficiently large N, where $\mathbb{1}$ is the indicator function. This is coherent with the intuitive definition of the invariant measure that we proposed. As analogs can be seen as recurrences around a given point, it appears here that the number of good analogs is highly linked to the invariant measure. This will be further discussed in Chapter II.

2   Analogs

2.1   Motivation and definitions

Analogs   An analog $\mathbf{a}$ of a state $\mathbf{x}$ is a state that is different but "close" to $\mathbf{x}$. In the following, $\mathbf{x}$ may be referred to as the *target state*. Defining what "close" means usually involves the use of a metric or distance function in the mathematical sense, although other types of measures of similarity could be used (this will be discussed next). The term

"analog" is used in atmospheric applications, while "k nearest neighbours" (kNN) is used in the machine-learning community. In analog forecasting —the principal analog method used in this thesis— one also defines *successors*. The successor is the state that follows the analog state. In this case, **x** is called the "present state", and the time-successors of **x**'s analogs are used to estimate the "future state".

Catalog   The catalog is the database in which the analogs are sought for. Analog methods are motivated by the ever-growing availability of geophysical datasets: observational data, output of numerical physical model, or reanalysis (a combination of observations and physical models through data assimilation). Those types of geophysical datasets correspond to one long trajectory of the system of interest, or multiple smaller trajectories. The aforementioned Poincaré theorem motivates the hope to find analogs of any given state. Note that these datasets are not necessarily in phase-space $\mathcal{P}$, especially for observations. Furthermore, the original dataset is often transformed to fit the needs of each application, such that the space in which the analogs lie (hereinafter referred to as "feature-space") may be different from the original observation space.

Feature space   The word "feature" is borrowed to the machine-learning community. Before applying analog methods, geophysical datasets are likely to be modified. This modification can be viewed as a mapping from observation-space to feature-space. A relevant feature-space is the main ingredient of analog methods. In practice, it has to informative and low-dimensional.

First, for a vector in feature-space to represent accurately the state of the system, the mapping from phase-space to feature-space must be injective. Observations are usually partial, such that the same observation can correspond to different states in phase-space. In this case, the use of time-embeddings is popular and motivated by the theorem of Takens (1981). It consists in concatenating the observation at time t and previous observations at times $t-\tau$, $t-2\tau$, etc., with fixed time-lag τ. Bounds for the minimum number of time-lags that allow the embedded space to be one-to-one with phase-space are known from theory (Sauer et al., 1991).

Second, analog methods are subject to the curse of dimensionality: the catalog size needed to find analogs of a given precision grows exponentially with feature-space dimension (see Van Den Dool 1994, Nicolis 1998 and the analysis of Chapter II). Therefore, dimension reduction techniques such as principal component analysis (also called em-

pirical orthogonal function) are commonly used. The practical choice of the dimension reduction technique is application-dependent, and this issue was not tackled during this PhD.

Examples and discussions on dimension reduction and Takens embeddings for analog forecasting can be found in Hamilton et al. (2016), Alexander et al. (2017), and Zhen et al. (2020). In this feature space, one must find an adequate measure of similarity between states.

Choice of distance   The Euclidean distance is often used to characterize "closeness". The properties of this distance are well known, and its popularity is such that algorithms for the search of nearest neighbours are often based on it. This allows for a simple, fast practical search of analogs.

Then, one strategy is to take the best analogs found using the Euclidean distance, and to measure their similarity with the target state $\mathbf{x}$ using other statistical quantities such as correlation or p-value, which are not sensitive to issues of low signal-to-noise ratio (Yiou et al., 2013). Similarly, Blanchet et al. (2018) use the Teweles-Wobus score to measure similarity in the shape (rather than actual value) of geopotential height fields.

The Euclidean-like Mahalanobis (1936) distance normalizes the data by their spatial covariance structure, which might be interesting. But this involves products of large matrices which also slows down the computation.

Fraedrich and Rückert (1998) use a Euclidean-like distance that adaptively gives more weights to certain coordinates of phase-space. In this case, the feature-space is made of time-embeddings of one coordinate of phase-space, and the weights are optimized for the task of analog forecasting. A similar procedure was used by Wetterhall et al. (2005) to select the best analogs of sea-level pressure fields projected on their first principal components, this time for the task of downscaling.

The Wasserstein distance finds the smallest path from one distribution to another, and is thus able to identify patterns that are similar up to a translation. This would allow to account for advection, for instance in meteorology, especially if the field is not smooth, in which case the Euclidean distance between a vector field and its advected counterpart might be large although they display similar physical features. It was used to detect changes in attractor structure of dynamical systems (Robin et al., 2017). However, this distance is computationally expansive, and is originally designed for probability distributions, which can complicate its use for real-valued vector fields (Thorpe et al., 2017).

Finally, additional conditions may be imposed for a state to be considered an analog of a given target state. For instance, the stochastic weather generator of Yiou (2014) selects analogs with a probability proportional to the calendar distance to the target state to ensure an averaged seasonal cycle in the simulated time series. Similar conditions may be imposed for the analogs to respect the seasonality or diurnal cycle of the target state.

2.2   Applications of analogs

Applications with a temporal component   Analogs of atmospheric circulation were introduced by Lorenz (1969) to study atmospheric predictability. Given a present state of the atmosphere $\mathbf{x}$ and its best analog $\mathbf{a}$, Lorenz (1969) uses the rate of divergence between the future states of $\mathbf{x}$ and the successors of $\mathbf{a}$ to evaluate the predictability of $\mathbf{x}$. More recently, analogs were often used in statistical schemes such as the stochastic weather generator of Yiou (2014) which produce several trajectories, allowing to asses the probability to observe a given atmospheric state. The use of analogs to generate ensemble forecasts was compared to other ensemble generating techniques by Delle Monache et al. (2013) and Atencia and Zawadzki (2017). Analog ensemble reforecasts allow Hamill and Whitaker (2006) to estimate rainfall cumulative probability distributions. Arroyo and Maté (2009) produce histogram forecasts of monthly precipitation at Chinese stations, using analogs of histograms based on the Wasserstein and Mallows distances to measure similarity. Analog forecasting was combined with data assimilation by Hamilton et al. (2016) and Lguensat et al. (2017), using either ensembles or Gaussian assumptions to estimate the distribution of analog forecasts. A direct application of analog forecasting with data assimilation is to fill gaps in observation maps, a method compared to optimal interpolation by (Zhen et al., 2020). Grooms (2020) combined analog forecasting with data assimilation and auto-encoders to compensate for an eventual lack of data (i.e., when the catalog is too small). Sévellec and Drijfhout (2018) discretize the possible values of global mean surface temperatures, allowing to perform fast forecasts with a transition matrix and analogs from CMIP5 trajectories. Combining analogs, Gaussian kernels and time-embeddings, Alexander et al. (2017) perform forecasts of tropical intraseasonal oscillations, based on a framework introduced by Zhao and Giannakis (2016). More recently, Wang et al. (2020) used analogs in a generalized, operator-theoretic framework (detailed by Alexander and Giannakis, 2020) to forecast the El Niño-Southern Oscillation.

Other applications Many applications of analogs do not involve time-evolution in the catalog (i.e., they use only analogs and not their successors). The most famous is analog downscaling where, given large-scale information, the possible states at a meteorological station (involving local, small-scale processes) are estimated. For instance, Wetterhall et al. (2005) produce statistics of precipitation at stations in Sweden from analogs of large-scale sea-level pressure fields and associated local precipitation values. Analogs also allow upscaling: from local information, the probability of large-scale patterns is estimated statistically in a similar fashion. This allows to reconstruct high-resolution fields at dates for which only spurious local observations are available (Schenk and Zorita, 2012; Yiou et al., 2013). Similarly, analogs can be used to specify a given variable from analogs of another variable (Tandeo et al., 2016). Finally, the role of atmospheric circulation in the occurrence of extreme events can be estimated using analogs. When observing an event such as a cold winter (Cattiaux et al., 2010), an increase in ice-sheet surface melt (Fettweis et al., 2013), or a heatwave (Jézéquel et al., 2018), one can find analogs of the respective atmospheric circulations and compare the associated temperatures, or ice melt. This allows to estimate the probability to observe extreme events assuming similar circulation patterns.

kNN and analogs The principle of k-nearest neighbours (kNN) is identical to the one of analogs. However, "kNN" usually refers to machine-learning applications for other types of datasets such as photographs, emails, traffic... For such datasets, the notions of dynamical systems, recurrences, and ergodicity are not necessary relevant, while they are central to the study of geophysical analogs. Therefore, we do not review applications of kNN here.

2.3 Strengths and weaknesses

An obvious advantage of using analogs is that it *circumvents the use of a physical model.* In particular physical models for the time evolution of specific observables (e.g., the ENSO index) are usually not provided or hard to write. Then, when a physical model is available, its numerical resolution might be numerically demanding. Furthermore, in such models, non-linear processes imply interactions between physical scales, which calls for the parametrization of unresolved processes at small scale (due to discretization). This results in uncertainties which might be hard to evaluate.

Analog methods have the advantage of being *computationally cheap*, such that it is *easy to perform ensemble forecasts*, allowing to assess for uncertainties. Machine-learning libraries allow for a fast search of analogs, with "trees" computed offline (see the appendix of Lguensat et al., 2017), and the application of analog methods usually consists in simple linear combinations. Contrarily, when using model-based methods from physical knowledge, one cannot hope to perform very large ensemble forecasts due to the high computational cost.

Analog methods have the advantage of being *non-parametric*. Which means that only a low number of meta-parameters has to be set in advance (the number of analogs used at each iteration, the choice of the distance, ...etc.). Other techniques have to estimate parameters, assuming a certain functional form or model. In particular, when one uses analogs in a statistical framework, there is no (necessary) assumption on the shape of probability distributions. This property is similar to the one of particle filters which will be introduced next. Note that analog methods can also be used in a parametric form when relevant: see the combination of analogs and Kalman filter of Lguensat et al. (2017) or the aforementioned weights optimization of Fraedrich and Rückert (1998) and Wetterhall et al. (2005).

Another strength of analog methods is that they *bear characteristics of the system* by construction. Atencia and Zawadzki (2017) noted that, contrarily to other initial condition perturbation methods, analog ensemble forecasts are consistent with the system's fractal dimension. Zhen et al. (2020) note that the analog data assimilation technique allows to build reconstructed sea-surface height maps with a temporal spectra that is more consistent with the true spectra then optimal interpolation. The random weather generator of Yiou (2014) preserves spatial constraints by construction, contrarily to other weather generators.

The simple principle behind analog methods makes them easily *interpretable*, while other data-based method might be seen as "black-boxes" where it is hard to understand why and when the method succeeds or fails. In the case of analog forecast, results of *convergence* in the limit of large data are reported in Zhao and Giannakis (2016) for analog forecasting with dynamics-adapted kernels, in the third chapter of this thesis and in Platzer et al. (2019) for analog forecasting in phase-space, and finally in Alexander and Giannakis (2020) for analog forecasting combined with kernel projection methods that are connected to Koopman operator theory.

Nowadays, numerical weather prediction is based on physical models combined with

observations through data assimilation (Carrassi et al., 2018). Yet, there was a time when analog forecasting was used for weather forecasting (Schuurmans and CJE, 1973). In a way, one could think that physical models have "won the battle". However, for the reasons mentioned earlier, analog methods are now used for their ability to perform statistical tasks such as downscaling, upscaling, event attribution, interpolation, or ensemble forecasts. Also, analog methods remain interesting for the forecast of specific observables (see Xavier and Goswami, 2007; Alexander et al., 2017; Wang et al., 2020; Ayet and Tandeo, 2018).

One of the weaknesses of analog methods is that they are subject to the *curse of dimensionality*: the number of data that is necessary for a given efficiency is an exponentially growing function of the dimension. This issue is studied in details in Chapter II, and consequences for analog forecasting are discussed in Chapter III. However, the curse of dimensionality is a generic concern for data-based methods. Furthermore, as stated before, it is now to adequately tune the feature-space in order to circumvent this issue.

A possible drawback of analog methods is that they are based on the hypothesis of *ergodicity*, such that they may not be able to generalize out of the previously observed states of the system. In particular, if the dynamics of the system change (e.g., due to global warming, change in land use...), analog forecasts are likely to be biased. However, as mentioned earlier, this drawback is turned into an advantage when concerned with the attribution of a given event to climate change. Note also that the formulation of Alexander and Giannakis (2020) remains well-posed in non-ergodic contexts.

3   Particle filtering

For an in-depth introduction to particle filters, the reader is referred to chapters on sequential Monte Carlo methods of the book by Cappé et al. (2006).

Particle filters belong to the field of data assimilation. Data assimilation techniques allow to estimate the present and future states of a system from partial and noisy observations combined with a model for the time-evolution of the system state. Model uncertainties are also taken into account.

The system state at (discrete) time t is modeled by a random variable X_t, from which observations are taken, also modeled by a random variable Y_t. Such a representation is called a "state-space model". The task of data assimilation (also called filtering) is to evaluate the probability distribution of X_t given all previous observations $Y_{1:t}$ from

time 1 to time t. Particle filters propose to estimate these probability distributions from a finite, but large number of "particles" that sample these distributions. They can be seen as ensemble methods, and are also called sequential Monte Carlo methods. Many Monte-Carlo filtering techniques (e.g., bootstrap filter, auxiliary particle filters) exist and their different flavours depend on the problem at hand (a priori knowledge, size of the problem). As pointed out in the review paper of Fearnhead and Künsch (2018), "Particle filters are completely general, but often suffer from sample depletion. Ensemble Kalman filters are more robust, but rely on Gaussian assumptions." In this PhD, we will leverage the simplest version of particle filtering.

Assume that one can generate samples of the distribution $d\mathbb{P}(x_t \mid x_{t-1})$ of X_t conditionally on X_{t-1} (this is the "model"), and of the distribution $d\mathbb{P}(y_t \mid x_t)$ of Y_t conditionally on X_t (this is the "observation"). Also assume that one has access to t observations $y_{1:t}$. Then, a particle filter estimates the probability distribution of the state X_t from an ensemble of n Dirac delta functions or "particles" $x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(n)}$. The simple particle filtering algorithm used in Chapter V is as follows:

- *Initialize* the particles $x_1^{(1)}, \ldots, x_1^{(n)}$ by taking n samples from an initial or "prior" distribution $d\mathbb{P}(x_1)$. The prior distribution is thus represented by a sum of dirac delta functions with equal weights $\omega^{(k)} = \frac{1}{n}$ (where $\omega^{(k)}$ designates the weight associated to the k-th particle).

- For each time $j < t$:

  - *Select* the particles that fit the observation at time j. This is done through an *update* of the particle weights according to Bayes' rule[6]. The weight of the k-th particle is multiplied by a factor $d\mathbb{P}(y_j \mid x_j^k)$:

    $$\omega^{(k)} \to \omega^{(k)} \frac{d\mathbb{P}(y_j \mid x_j^k)}{\sum_{l=1}^n d\mathbb{P}(y_j \mid x_j^l)}$$

    Then a new set of n particles $\tilde{x}_j^{1:n}$ is drawn from the previous set of particles with updated weights. Thus, the particles for which the observation y_j is "credible" are likely to be drawn multiple times, while other particles might not be drawn. The resulting *updated* set of particles give an approximation of the distribution of X_j conditionally on $Y_{1:j} = y_{1:j}$.

---

6. Here we show the most popular weight updating formula, called "sequential importance resampling", but there are other strategies.

– *Predict* the future state by sampling each particle x_{j+1}^k from the probability distribution $\mathrm{d}\mathbb{P}(x_{j+1} \mid \tilde{x}_j^k)$. The set of particles $x_{j+1}^{1:n}$ gives an approximation of the distribution of X_{j+1} given that $Y_{1:j} = y_{1:j}$.

One advantage of particle filters is that they are nonparametric: there is no assumption on the shape of the distribution of X_t and Y_t. Therefore, particle filters are able to model any kind of distribution. Parametric data assimilation methods include Kalman filters, which assume Gaussian random variables. A major drawback of particle filters is that in high dimension, one needs a very large number of particles to sample whole probability distributions.

4 Extreme value theory

For a general introduction to extreme value theory, the reader is referred to the book of Coles et al. (2001).

Extreme value theory (EVT) is devoted to the study of the probability for a random variable to reach high values. One important task of EVT is to extrapolate return levels (quantiles) beyond the largest sample value, and to infer the confidence intervals of such high quantile estimates. There are two main approaches in EVT: block-maxima and peak-over-threshold (POT). The first one studies the limiting distribution for the maxima of fixed-length samples of i.i.d. random variables, when the sample length goes to infinity. The second one gives the limiting distribution of the probability for a random variable to exceed a threshold when this threshold is infinitely large.

Let a real-valued random variable Y, from which are taken samples of length n. Denote M_n the maximum of these samples. If, after rescaling with series a_n and b_n, the cumulative distribution function of $(M_n - b_n)/a_n$ converges to a non-degenerate distribution function G when $n \to \infty$, then G has to belong to the family of generalized extreme value (GEV) distributions :

$$G(y) = \exp\left\{ -\left[1 + \xi\left(\frac{y-\mu}{\sigma}\right) \right]^{-1/\xi} \right\},$$

where $\xi \in \mathbb{R}$, $\mu \in \mathbb{R}$ and $\sigma > 0$ are parameters that have to be estimated, and the function G is defined where $1 + \xi\left(\frac{y-\mu}{\sigma}\right) > 0$. The values of σ, μ, and ξ respectively control the scale, location and shape of the tail of the distribution of Y. Positive shape parameters ξ correspond to functions of the Fréchet family. Weibull distributions are found fo $\xi < 0$.

The case $\xi = 0$ is obtained by taking the limit $\xi \to 0$ in the equation defining G, and gives distributions of the Gumbel family.

This approach allows to determine the probability distribution of analog-to-target distances in Chapter II. In particular, we use the fact that if the maximum sample of Y is GEV-distributed as stated above, the k-th maxima follows the distribution G_k:

$$G_k(y) = \exp\left\{-\tau(y)\right\} \sum_{s=0}^{k-1} \frac{\tau(y)^s}{s!} \, ,$$

where $s!$ is factorial s and $\tau(y) = \left[1 + \xi\left(\frac{y-\mu}{\sigma}\right)\right]^{-1/\xi}$. Similar formulas for the joint probability distribution of the k first maxima are used to give the joint distirbution of the k-best analog-to-target distances.

The POT approach looks at the distribution of threshold exceedances. If the sample maxima of Y are GEV-distributed as above, then for large enough u:

$$\mathbb{P}\left(Y - u > v \,|\, Y > u\right) \approx \left(1 + \frac{\xi v}{\tilde{\sigma}}\right)^{-1/\xi} \, ,$$

where $\tilde{\sigma} = \sigma + \xi(u_\mu)$ and the function is defined for every $v > 0$ and $1 + \xi v/\tilde{\sigma} > 0$. This last distribution is called the generalized Pareto distribution (GPD), and is used to model the behaviour of heavy-tailed random variables in Chapter V.

EVT is an active research field where statistical models dealing with non-stationarities in space and time, as well as multivariate settings have been studied, see e.g. Davison and Huser (2015). In this dissertation, we will mainly borrow univariate EVT tools and concepts in an i.i.d. or stationary setup.

5 Ocean waves

For a deeper introduction to the physics of ocean surface gravity waves, the reader is referred to the books of Kinsman (1965).

5.1 Generalities

The separation between the ocean and the atmosphere can be seen as a moving surface, characterized by an elevation that depends on two horizontal coordinates and on time. In the absence of perturbations such as wind stress, earthquakes, or the moon's gravitational

attraction, the surface at rest is flat (for small enough horizontal distances, i.e. neglecting the curvature of the earth surface). Any perturbation of this flat surface is pulled back by the gravity force or by capillary forces (surface tension), which are responsible for the propagation of surface waves. Capillary waves or "ripples" are very small waves, and in this thesis we focus on the larger, *gravity waves.*

A surface wave is characterized by its wavelength λ (the distance between two crests or troughs), its wave period T (the time difference between two consecutive maxima at a given space location), its height H (either crest height or crest-to-trough wave height), and its direction.

In this thesis, we focus on wind-generated ocean waves. Other types of ocean waves include tsunamis (displacements of large bodies of water due to a movement of the earth surface), tides (due to gravitational forces of the moon and the sun), and slam waves (due to the fall of a large object on the water surface).

The generation of ocean waves from the wind initially generates waves of very small height and wavelength, and large waves are only generated if the wind blows strong enough, for a long enough time, and over a large enough distance (called the "fetch"). Such conditions are met in severe storms. These large amplitude waves then propagate due to earth's gravitational force, travelling over distances some times exceeding thousands of kilometers.

Waves of large wavelength travel faster then waves of small wavelength. This dispersion effect causes far-traveling waves of different wavelength to separate. Therefore, while the waves that have been generated by local wind (called *wind waves*) can look disordered, waves that have been generated far away (called *swell*) appear to be more regular and smooth. A wind-wave sea-state displays a wide range of directions, wavelengths and heights, while swell sea-states are generally organised around a given direction, wavelength and height. For instance, the waves that are surfed at the beach are generally swell waves.

In chapter IV, we focus on *swell-dominated sea-states.* In the next sections, we introduce the physics of ocean gravity waves *propagation*, and wave-wave interactions. We do not tackle the issues of wave-current interactions, nor wave-wind interactions, nor wave dissipation, nor the influence of sea-bottom on waves.

5.2 Linear waves and wave groups

The motion of gravity waves can be described assuming incompressible and inviscid water and wind, irrotational water, and *small steepness* of the waves (i.e., $H/\lambda \ll 1$). From

these physical assumptions, Navier-Stokes equations allow to derive the relation between the wave period and wavelength of a sinusoidal gravity wave, called the *dispersion relation* of linear gravity waves (we will come back to the term "linear" later). Expressed as a function of wavenumber $k = 2\pi/\lambda$ and pulse $\omega = 2\pi/T$, where λ is the wavelength and T the wave period, this dispersion relation reads:

$$\omega^2 = gk \,,$$

where g is the earth's gravitational acceleration at the ocean surface. The velocity of this sinusoidal wave is given by the ratio ω/k, and is thus a growing function of λ as stated earlier ($\propto \sqrt{\lambda}$).

This description is valid at *first order* in wave steepness kH, where H is the wave height. This approximation allows to describe the so-called *linear waves*, because the equations of wave motion have been linearized in the limit $kH \to 0$. Non-linear effects arise from solutions to the same physical equations, but for higher orders of expansion in wave steepness.

There is no interaction between linear waves of different wavelengths: they simply "cross" each other, and the total surface elevation is the sum of the contributions of each wave. Therefore, an ocean surface of sufficiently small steepness can be described as an infinite (or large) sum of sinusoidal waves that propagate according to the dispersion relation of linear gravity waves. From a statistical point of view, the probability distribution of the elevation of such a sea-surface is Gaussian (see the classical monograph of Longuet-Higgins, 1957). Therefore, this type of sea-surface is called a *Gaussian sea.*

This type of surface is therefore usually analyzed using Fourier decompositions. The "wave spectrum" is obtained from the square amplitude of the elevation's Fourier transform. The integral of the wave spectrum gives the total wave energy. The latter remains constant in the absence of energy input (from wind, or other sources listed above) and wave energy dissipation. The range of frequencies (or wavenumbers) over which the spectrum is significantly high defines the spectral bandwidth. From the previous chapter, it follows that swell seas are *narrow-banded* and wind seas are *wide-banded.*

The hypothesis of narrow spectrum allows for several approximations, including the description of waves using *wave groups* or *wave packets.* It is well known that that the sum of two sinusoïds with similar frequencies can be viewed as the product of a sinusoïd with the same frequency as the original waves (called the *carrier wave*) and a low-frequency amplitude or *envelope.* The same can be said of a large (or infinite) sum of sinusoïds with

similar frequencies. Thus, swell seas are organized in groups, a phenomenon that surfers know well: a large swell wave is usually followed by several other large swell waves.

5.3 Non-linear waves

Solving the same physical equations as for linear waves, but relaxing a little the hypothesis of small steepness (i.e. $kH \rightarrow 0$), one can find new solutions that correspond to expansions with respect to the small parameter kH.

The solution including the next expansion term after linear waves is called *second-order* non-linear waves. In this solution, two waves of wavenumbers k_1 and k_2 can interact to create waves with wavenumbers equal to the sum and difference of original wavenumbers $k_1 + k_2$ and $|k_1 - k_2|$. The first created wave has a high frequency and thus changes the shape of waves, while the second has a low frequency and thus changes local mean levels of sea-surface elevation (see, for instance, the figures of Forristall, 2000). In the case of an initially individual, sinusoïdal, linear wave of wavenumber k, second-order non-linearities add a high-frequency wave of wavenumber $2k$, which travels at the same velocity as the initial linear wave. The resulting wave is called the Stokes wave, and has sharper crests and wider troughs than the linear wave. These added waves that are the consequence of second-order, *two-wave interactions*, travel at the same velocity then the linear waves they originate from, and are therefore termed *bound waves*. Consequently, the latter *do not respect the dispersion relation* of linear waves. Another important aspect of second-order wave physics is that, as in the linear case, there is no exchange of energy between waves of different wavelengths. The wave-spectrum thus remains time-invariant, which is not the case for higher-order non-linear ocean waves.

In the following, the term *higher-order non-linear waves* designates solutions of the same wave-physics equations as above, but including terms in the wave-steepness expansion that are higher than second-order terms. These include three-wave and four-wave interactions. Such higher-order non-linear waves exhibit exchanges of energy between different wavelengths, and therefore the spectrum of such waves varies in time, even when there is no energy input or dissipation. A consequence of these interactions is the Benjamin-Feir (1967) instability, which causes any individual linear wave solution to be unstable to side-band perturbations. This mechanism may be responsible for the formation of waves of large height. The non-linear Schrödinger equation (NLSE) provides a narrow-band approximation of the time evolution of the wave envelope. The NLSE is valid at third order in wave steepness, and is therefore said to correspond to "weakly

non-linear" ocean waves. This equation is used in Chapter IV to evaluate the effect of higher-order non-linearities on our prediction method.

5.4   Rogue waves

In order to go beyond this brief introduction, the reader is referred to the numerous literature reviews on the topic of oceanic rogue waves (Kharif and Pelinovsky, 2003; Müller et al., 2005; Dysthe et al., 2008; Adcock and Taylor, 2014; Slunyaev, 2017).

The concept of rogue wave emerges from sailors' stories. From the "mythical" point of view, a wave is rogue when it emerges from nowhere, disappears in the same fashion, and is abnormally large compared to its neighbourgs[7]. Rogue waves were consider a sailors' tale until scientific research revealed that waves with such properties can be measured in the ocean, reproduced in laboratory and numerical experiments, and correspond to some analytical solutions of physical equations.

Numerous offshore accidents have been attributed to rogue waves. The hazardous potential of rogue waves is a motivation for research on this topic.

A quantitative definition has been retained by the scientific community, saying that a wave is "rogue" if its crest-to-trough wave height is at least twice as large as the significant wave height:

$$H \; > \; 2\,H_s\,, \tag{I.1}$$

where the significant wave height H_s can be defined as four times the standard deviation of the sea-surface elevation (statistics computed from samples which are assumed to originate from a constant wave spectrum, which may involve practical issues). Note that the exact threshold is arbitrary. It correspond to relatively rarely observed waves. Similar criteria are based on crest height rather than crest-to-trough wave height. Assuming linear waves with Gaussian probability distribution and a relatively narrow spectrum, one finds that one wave out of 10^4 is rogue according to the definition of $H > 2H_s$. The average number of encountered rogue waves varies when including other types of wave physics.

Several physical phenomenon are candidates to explain the occurrence of oceanic rogue waves. They include the linear superposition of a large number of ocean waves or "linear focusing", the crossing of sea-states from different directions, the aforementioned

---

7. We retain this description, although other the term "rogue wave" may refer to other wave-phenomenon when used by non-physicists.

Benjamin-Feir instability, wave-current interactions, wind blowing over high amplitude waves, and the influence of bathymetry on waves.

In Chapter IV, we use the linear approximation of ocean waves. Due to the narrow-banded character of swell waves, we assume that the surface can be studied from the evolution of isolated wave packets. The effects of higher-order non-linearities are studied through the non-linear-Schrödinger equation.

How near are the best analogs?

As around the sun the earth knows she's revolving
And the rosebuds know to bloom in early May...

<div align="right">

– Stevie Wonder, *As*

</div>

1 Introducing the article

How near are the nearest past states of today's weather? This questions is not new, although the formulation and the point of view taken in this study are different from past research work on the matter. The distance between the target state (say, today's weather) and the best analogs taken from a catalog (say, a historical database) is a fundamental quantity for any application of analog methods, including analog forecasting. An underlying hypothesis of analog methods is that the analogs are "good" in the sense that they are close to their target. Knowing a priori how close the analogs will be allows one to estimate the performances of a given method before using it. This chapter investigates the following questions:

- Given a catalog size and a target point of a dynamical system, how far from this target will the catalog's best analogs be? What is the variability of analog-to-target distances?

- What is the influence of catalog size and dimension on the nearness of analogs?

To answer these questions, the following strategies and tools are used:

* The analogs are viewed as minimizers of a distance map on the dynamical system's attractor. In the limit of large catalog size, the tools of extreme value theory can be applied.

* Analytical probability distributions for analog-to-target distances are derived based
  on strong regularity hypothesis for the invariant measure of the flow. Those dis-
  tributions yield a simple form. Their shape and scale are functions of the local
  dimension and catalog size.

* It is the first time that these distributions are proposed to analyze the link between
  analog performances, catalog size, and dimension. Previous studies focused on re-
  turn times and were limited to the first analog. This change of point of view brings
  new insights for applications of analog methods.

* The applicability of these distributions is examined in numerical experiments of the
  Lorenz (1963) system, and using 10-m wind data from AROME model output.

2 Article in preparation for *Entropy*: "Probability distributions for analog-to-target distances"

Paul Platzer[1,2,3,4] ● Pascal Yiou[1,2] ● Philippe Naveau[1,2] ● Pierre Tandeo[3] ●
Jean-François Filipot[4]

2.1 Abstract

Analogs are nearest neighbors of a target state. They are taken from a large database
called the catalog. Analogs have been used in many atmospheric applications including,
among others, forecasts, downscaling, and predictability estimation. The distances of
the analogs to the target state largely condition the performances of analog applications.
These distances are random variables, and their probability distributions can be related
to the catalog size and properties of the system at stake. At the moment, authors have
focused on the first moments of return time statistics for the best analog, fixing an ob-
jective distance from this analog to the target state. However, for practical use and to
reduce estimation variance, applications usually require not just one, but many analogs.
In this paper, we evaluate theoretically and with numerical experiments the probability

1. LSCE, CEA Saclay l'Orme des Merisiers, UMR 8212 CEA-CNRS-UVSQ, U Paris-Saclay, IPSL,
Gif-sur-Yvette, France
2. ESTIMR - Extrèmes : Statistiques,Impacts et Régionalisation, LSCE, Gif-sur-Yvette, France
3. IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238, Plouzané, France
4. France Énergies Marines, Plouzané, France

distributions of the K-best analog-to-target distances. We show that dimensionality plays a role not only in the size of the catalog needed to find good analogs, but also in the relative means and variances of the K-best analogs. Our results are based on well-known tools from dynamical systems theory. These findings are confirmed with numerical simulations of famous chaotic dynamical systems and on 10m-wind numerical model output in north-west France.

2.2 Introduction

Atmospheric analogs have been introduced by Lorenz (1969) in a study on atmospheric predictability. The faster one target state z and its best analog a_0 diverge from one another, the harder it is to predict the evolution of z. In Lorenz's study, the state z was characterized by height values of the 200-, 500- and 850-mb isobaric surfaces at a grid of 1003 points over the Northern Hemisphere. The database of available analogs, called the catalog, contained five years of twice daily values. In his abstract, Lorenz states that there are "numerous mediocre analogues but no truly good ones".

Since Lorenz's work, analogs have been used in many applications such as weather generators (Yiou, 2014), data assimilation (Hamilton et al., 2016; Lguensat et al., 2017), kernel forecasting (Alexander et al., 2017), downscaling (Wetterhall et al., 2005) climate reconstruction (Schenk and Zorita, 2012; Yiou et al., 2013) and extreme event attribution (Cattiaux et al., 2010; Fettweis et al., 2013; Jézéquel et al., 2018).

The reason why Lorenz could not find any good analog was made clear later on by Van Den Dool (1994). It was shown that for high-dimensional systems, the mean recurrence time of a good analog (identified as a minimum catalog size) grows exponentially with dimension. This result is a variant for analogs of the "curse of dimensionality", well known in data-sciences. With three pressure levels over the whole Northern Hemisphere, the dimension of Lorenz's study was very high, and only five years of twice-daily data was not enough to ever hope to find a good analog.

Nicolis (1998) added a dynamical system's perspective to Van den Dool's analysis. She showed that studying mean recurrence times was not enough, as the relative standard deviation of this recurrence time could be very high. Furthermore, it was shown that recurrence time statistics exhibit strong local variations in phase-space, so that certain target states may need a larger catalog size to find good analogs.

Accounting for Van den Dool's findings, it is now usual to reduce as much as possible the feature-space dimension before searching for analogs. Also, the last decades have

witnessed a proliferation of data from in-situ and satellite observations, as well outputs from numerical physics-based model. Such conditions allows one to find good analogs in many situations, and it has become standard to use not just one, but many analogs (usually a few tens). From a statistical perspective, using many analogs instead of one can increase estimation bias, but it reduces estimation variance, so that the estimation is less sensitive to noise. Using many analogs also allows to perform local regression techniques on the analogs, such as local linear regression (Lguensat et al., 2017). This technique has proven efficient in analog forecasting applications (Ayet and Tandeo, 2018), and it was shown that local linear regression allows analog forecasting to capture the local Jacobian of the dynamics of the real system (Platzer et al., 2020, under revision).

This new context suggests to focus not only on the best analog a_0, but also the k-th best analog, for k up to $\sim$40. Also, one can now reasonably hope to find good analogs using dimension reduction and a large amount of data. Thus, one is less interested in recurrence-times, but rather in analog performances. Performances of analog-based methods are largely conditioned by analog-to-target distances. In this work we propose to evaluate the probability distribution of these distances. Our analytical probability distributions make the link between analog-to-target distances, catalog size and local dimension. This brings new insight on the impact of dimensionality on analog forecasting performances.

The first section outlines the theoretical framework and findings. The second section interprets the findings and compares this analysis with past studies. The third section shows results from numerical experiments of the Lorenz (1963) system and from 10-m wind data from the numerical model AROME.

2.3 Theory

2.3.1 Analogs in dynamical systems and local dimensions

In ergodic systems, trajectories come back infinitely close to their initial condition after a sufficiently long time (Poincaré, 1890). Furthermore, if the dynamical system has an attractor set $\mathcal{A}$, then all trajectories converge to this subset of the phase-space (Milnor, 1985). Analog methods are based on the idea that if one is provided with a long enough trajectory of the system of interest, one will find analog states close to any point z of the attractor $\mathcal{A}$.

The trajectory from which the analogs are taken is called the "catalog" $\mathcal{C}$, and can

either come from numerical model output or reprocessed observational data. It can be seen either as a trajectory from a discrete dynamical system, or as evenly-spaced time samples from a continuous dynamical system. In any case, the catalog has a finite number of elements noted $L := \mathbf{card}(\mathcal{C})$. This catalog size may be divided by a typical correlation time-scale so that elements of the catalog can be considered independent (Van Den Dool, 1994). In fact, for the analogs of a given target z to be considered independent, it is enough that the typical distance between two analogs of z be smaller than the typical distance between an analog and its time-successor.

The structure of the attractor, expressed by the systems' invariant measure μ, conditions the structure of the catalog and the ability to find analogs. In particular, Van Den Dool (1994) and Nicolis (1998) studied the role of the attractor's dimension. Let $B_{z,r}$ the ball centered on $z \in \mathcal{A}$ and of radius r, then

$$d_{z,r} := \frac{\log \mu(B_{z,r})}{\log r}, \tag{II.1}$$

defines the finite-resolution (r-resolution) local dimension at point z. Note that for ergodic measures, $\mu(B_{z,r})$ can be approximated by counting the number of times a given trajectory enters $B_{z,r}$ (this is the consequence of the ergodic theorem of Birkhoff 1931).

If μ is ergodic and $\lim_{r \to 0} d_{z,r}$ exists, then μ is said to be exact dimensional and the limit is independent of z (Young, 1982). This typical value of the local dimension is noted D_1.

$$D_1 := \lim_{r \to 0} d_{z,r}.$$

The finite-resolution local-dimension $d_{z,r}$, however, can deviate from the typical value D_1. More precisely, $d_{z,r}$ exhibits large deviations from its limit value (for more details, see Caby et al., 2019).

The distance from the k-th analog $a_k \in \mathcal{C}$ to the target state z is noted $r_k := \mathrm{dist}(a_k, z)$. Distances are sorted so that $r_1 < r_2 < \ldots < r_K$, and K is the total number of analogs considered. Empirical methods usually set K to a fixed value, reaching for a bias-variance trade-off. This amounts to looking at a lower quantile of the function $x \to \mathrm{dist}(x, z)$. Another possibility is to set a threshold R for the analog-to-target distances so that $r_K < R < r_{K+1}$. In this case, K depends on z.
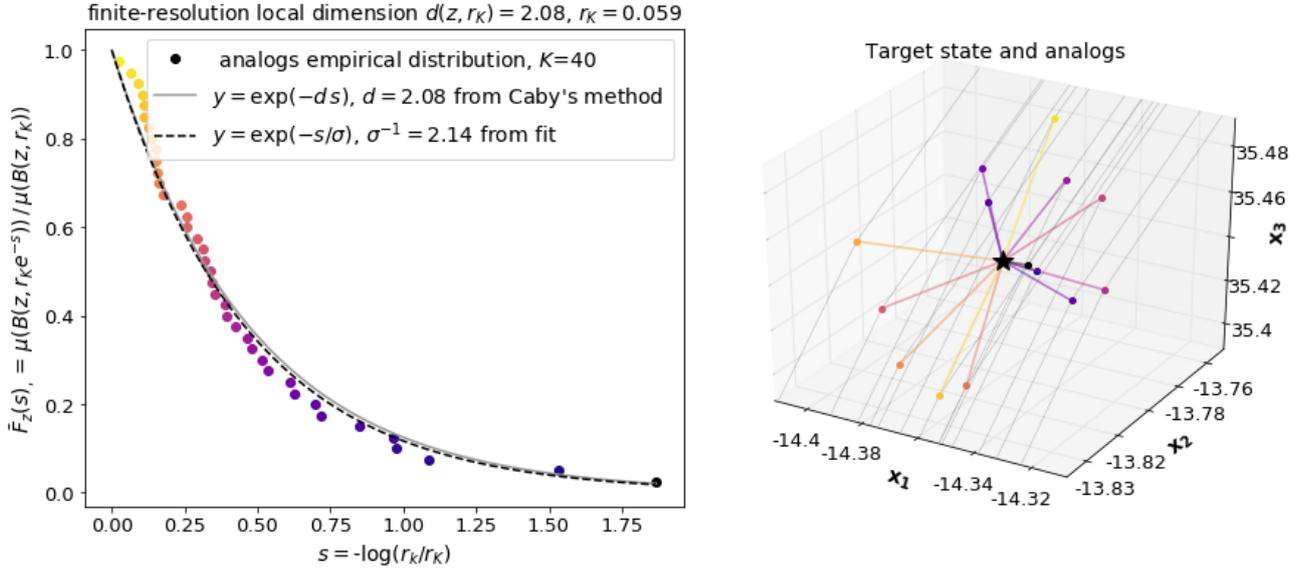
Figure II.1: Computing the finite-resolution local dimension d_{z,r_K} at a point z of the three-variable Lorenz (1963) system. Following from Caby et al. (2019), we evaluate d by taking the mean of the empirical cumulative distribution function shown in left panel. For this example, fitting the empirical CDF with an exponential $\exp(-s/\sigma)$ and taking the inverse of σ would have given approximately the same value for d. The right panel shows the target z as a black star, and (one in three) analogs as colored dots (colors match the left panel). The trajectories from which the analogs are taken are in gray. In this example, the typical analog-to-successor distance is much larger than the typical analog-to-target distance.

2.3.2 Simple scaling of analog-to-target distance with dimension

Using extreme value theory and dynamical systems theory, Caby et al. (2019) showed that $d_{z,r}$ can be estimated using the empirical distribution of points inside a ball of exponentially decreasing radius. This empirical distribution is actually exactly the cumulative distribution function of the best available analogs. It then follows from Caby et al. (2019) that, for regular enough measures, we have the approximate scaling:

$$r_k(z) \sim k^{1/d}, \tag{II.2}$$

where $d = d_{z,r_K}$ is the local dimension at finite resolution r_K (the largest analog-to-target distance). An application of this method to the three-variable system of Lorenz (1963) is given in Fig. II.1.

Eq. II.2 gives an important point of our analysis, which is the scaling of r_k with k,

and is approximately given by a power-law with exponent $1/d$. However, this formula comes from a work on local dimensions, not analog-to-target distances. It is therefore not surprising that some of the elements required for our study are missing. In particular, this scaling does not give the constant in front of $k^{1/d}$, in which resides the relation to the catalog size, a crucial point for analog applications. Also, it only gives a mean or typical value of r_k, while our objective is to evaluate the probability distribution of r_k, or at least the probability of departures from this mean scaling.

The next section gives theoretical elements to evaluate the full probability distribution of $r_k(z)$ from the local dimension, the catalog size, and the analog number k.

2.3.3   Full probability distribution of analog-to-target distance

Poisson distribution of the number of analogs in a ball   Haydn and Vaienti (2019) have shown that, for dynamical systems having Rare Event Perron-Frobenius Operator properties, and for non-periodic points z, the number of visits $k(z, r)$ of a trajectory of size L into the ball $B_{z,r}$ follows a Poisson distribution with mean $L\mu(B_{z,r})$:

$$\mathbb{P}\left(k(z,r) = k\right) = \frac{\left(L\mu(B_{z,r})\right)^k}{k!} e^{-L\mu(B_{z,r})}, \tag{II.3}$$

where $k!$ is k factorial. In the context of analogs, this is the probability to find k analogs with distances to z below the radius r. In the following we write $\mu_{z,r} := \mu(B_{z,r})$.

Distribution of analogs close to the sphere   This section aims at using μ to evaluate $\mathbb{P}\left(r_k \in [r, r+\delta r)\right)$, the probability that the k-th analog-to-target distance is between r and $r + \delta r$, for fixed k and z and where δr is small compared to r.

The event "$r_k \in [r, r+\delta r)$" is the intersection of the event "there are $k-1$ analogs in the ball $B_{z,r}$" and the event "there is one analog in $B_{z,r+\delta r} \cap \overline{B_{z,r}}$". For a Poisson point process these two events are independent, so that:

$$\begin{aligned} \mathbb{P}\left(r_k \in [r, r+\delta r)\right) &= \mathbb{P}\left(k(z,r) = k-1 \ \wedge \ \exists x \in \mathcal{C} \cap B_{z,r+\delta r} \cap \overline{B_{z,r}}\right) \\ &= \mathbb{P}\left(k(z,r) = k-1\right) \mathbb{P}\left(\exists x \in \mathcal{C} \cap B_{z,r+\delta r} \cap \overline{B_{z,r}}\right) \\ &= \frac{\left(L\mu_{z,r}\right)^{k-1}}{(k-1)!} e^{-L\mu_{z,r}} \mathbb{P}\left(\exists x \in \mathcal{C} \cap B_{z,r+\delta r} \cap \overline{B_{z,r}}\right). \end{aligned} \tag{II.4}$$

Then, it follows from Haydn and Vaienti (2019) that the event that strictly one element

of the catalog lies between $B_{z,r}$ and $B_{z,r+\delta r}$ has a probability of the same form as Eq. (II.3) but replacing k by 1 and $\mu_{z,r}$ by $\delta\mu_{z,r} := \mu_{z,r+\delta r} - \mu_{z,r}$

$$\mathbb{P}\left(\exists! x \in \mathcal{C} \cap B_{z,r+\delta r} \cap \overline{B_{z,r}}\right) = L\delta\mu_{z,r}e^{-L\delta\mu_{z,r}}\,. \tag{II.5}$$

If the invariant measure μ is regular enough so that $\lim_{\delta r \to 0}\delta\mu_{z,r} = 0$ we then have $e^{-L\delta\mu_{z,r}} \approx 1$. Also, the probability to find more than one element of the catalog between $B_{z,r}$ and $B_{z,r+\delta r}$ has a probability of $\mathcal{O}(\delta\mu_{z,r})^2$. This justifies the approximation $\mathbb{P}\left(\exists x \in \mathcal{C} \cap B_{z,r+\delta r} \cap \overline{B_{z,r}}\right) \approx \mathbb{P}\left(\exists! x \in \mathcal{C} \cap B_{z,r+\delta r} \cap \overline{B_{z,r}}\right)$. Finally, combining Eq. (II.4) and Eq. (II.5), one finds:

$$\mathbb{P}\left(r_k \in [r, r+\delta r)\right) = L\delta\mu_{z,r}\frac{\left(L\mu_{z,r}\right)^{k-1}}{(k-1)!}e^{-L\mu_{z,r}} \tag{II.6}$$

This last equation is a more general form of our main result which is given in the next section. Here, the probability is expressed in terms of the invariant measure, which is usually not known analytically. The next section expresses the same probability in terms of the analog-to-target distance r.

Distribution of analogs-to-target distance The link between $\mu_{z,r}$ and r is given by the definition of the finite-resolution local dimension in Eq. (II.1)

$$\mu_{z,r} = r^d\,, \tag{II.7}$$

where $d = d_{z,r}$. The link between $\delta\mu_{z,r}$ and δr involves variations of the local dimension wih r. Let $\Delta = d_{z,r+\delta r} - d_{z,r}$, we have:

$$\frac{\delta\mu_{z,r}}{\mu_{z,r}} = \left(1 + \frac{\delta r}{r}\right)^{d+\Delta}e^{\Delta\log r} - 1\,. \tag{II.8}$$

Using the regularity hypothesis $\Delta \ll d$, and keeping only lower-order terms, we find:

$$\frac{\delta\mu_{z,r}}{\mu_{z,r}} = d\frac{\delta r}{r} + \Delta\log r\,. \tag{II.9}$$

The term $d\frac{\delta r}{r}$ represents an almost steady increase in $\mu_{z,r}$ when r grows. The term $\Delta\log r$ represents fluctuations in this increase given by the fluctuations in $d_{z,r}$. In practice, the method described in Sec. 2.3.2 to evaluate d should catch a mean local dimension

54

over the analogs and not catch the fluctuations of $d_{z,r}$ with r at scales smaller than r_K. Thus, the approximation :

$$\frac{\delta \mu_{z,r}}{\mu_{z,r}} \approx d \, \frac{\delta r}{r} \, , \tag{II.10}$$

which is not valid in theory, should be relevant in practice for finite catalog size and regular enough measures. For small enough δr, one can then define p_k, the probability density function of r_k through the identity $\mathbb{P}\left(r_k \in [r, r+\delta r)\right) = p_k(r)\delta r$. Combining Eq. (II.6), Eq. (II.7) and Eq. (II.10), we find:

$$p_k(r) = d \, L \, r^{d-1} \, \frac{\left(L \, r^d\right)^{k-1}}{(k-1)!} \, e^{-L \, r^d} \, . \tag{II.11}$$

This last equation is our main result. An alternative proof for Eq. (II.11) using Extreme Value Theory is given in Sec. 2.3.5. Eq. (II.11) then allows to compute the mean and variance of r_k for fixed k and d :

$$\langle r_k \rangle = \frac{\Gamma\left(k + \frac{1}{d}\right)}{L^{1/d} \, \Gamma(k)} \, , \tag{II.12a}$$

$$\langle r_k^2 \rangle - \langle r_k \rangle^2 = \frac{1}{L^{2/d} \, \Gamma(k)^2} \left\{ \Gamma\left(k + \frac{2}{d}\right)\Gamma(k) - \Gamma\left(k + \frac{1}{d}\right)^2 \right\} \, , \tag{II.12b}$$

where Γ is Euler's Gamma function. These identities can be simplified through scalings of the Gamma function $\Gamma(x+1) = \int_0^{+\infty} u^x e^{-u} \mathrm{d}u$ for large x, using Laplace's method up to second order to evaluate the integral (the first order gives Stirling's formula). This gives:

$$k \geq 2 \, , \, \langle r_k \rangle \approx \left(\frac{k}{L}\right)^{1/d} \, , \tag{II.13a}$$

$$\frac{\left(\langle r_k^2 \rangle - \langle r_k \rangle^2\right)^{1/2}}{\langle r_k \rangle} \approx \frac{1}{dk^{1/2}} \, , \tag{II.13b}$$

where we find again the scaling $r_k \sim k^{1/d}$ of Eq. (II.2). These approximations will be increasingly valid as k grows, but even for $k = 2$, Eqs. (II.13a,b) give a satisfactory numerical approximation of Eqs. (II.12a,b).

One can also compute r_k^*, the value of r for which p_k reaches a maximum:

$$r_k^* = \mathrm{argmax}_r \left\{ p_k(r) \right\} = \left(\frac{k - \frac{1}{d}}{L} \right)^{1/d} ,$$

and when $kd \leq 1$, $r_k^* = 0$ and $p_k(0) = +\infty$. Note that the three quantities $\langle r_k \rangle$, $\left(\frac{k}{L} \right)^{1/d}$ and r_k^* are equivalent as $k \to +\infty$.

Fig. II.2 shows plots of $p_k(r)$ against r for varying values of d and k. As a consequence of the scaling $r_k \sim k^{1/d}$, we observe large variations of $\langle r_k \rangle$ with k for small dimensions d, and very small variations of $\langle r_k \rangle$ with k for large dimensions d. Note that, in the limiting case $d \to \infty$, the random variables r_k are degenerate and all equal $L^{-1/d}$ almost surely. This can be witnessed through the different scales of the horizontal axis of the plots. Also, as a consequence of Eqs. (II.13), we have that the standard deviation of r_k is a growing function of k for $d < 2$, while it is constant for $d = 2$ and decreasing for $d > 2$. However, the relative standard deviation of r_k is always a decreasing function of k and d according to Eq. (II.13b).

2.3.4   Rescaling and convergence to the standard Normal distribution

Eqs. (II.13a,b) suggest the change of variables from r to u with

$$u = dk^{\frac{1}{2}} \left(\left(\frac{L}{k} \right)^{\frac{1}{d}} r - 1 \right) ,$$

so that the probability density function of u_k, noted $h_k(u)$, is

$$h_k(u) = \frac{k^{k-\frac{1}{2}}}{(k-1)!} \left(1 + \frac{u}{dk^{\frac{1}{2}}} \right)^{dk-1} \exp \left\{ -k \left(1 + \frac{u}{dk^{\frac{1}{2}}} \right) \right\} , \qquad \text{(II.14)}$$

and simple asymptotic analysis gives

$$\lim_{k \to +\infty} h_k(u) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{u^2}{2} \right\} ,$$

which shows that the rescaled random variable u_k converges in distribution to the standard Normal distribution as $k \to +\infty$. Note, however, that this limit should be hard to observe in practice, as the distribution of Eq. (II.11) is valid only in the limit of large catalog size and with $k \ll L$.
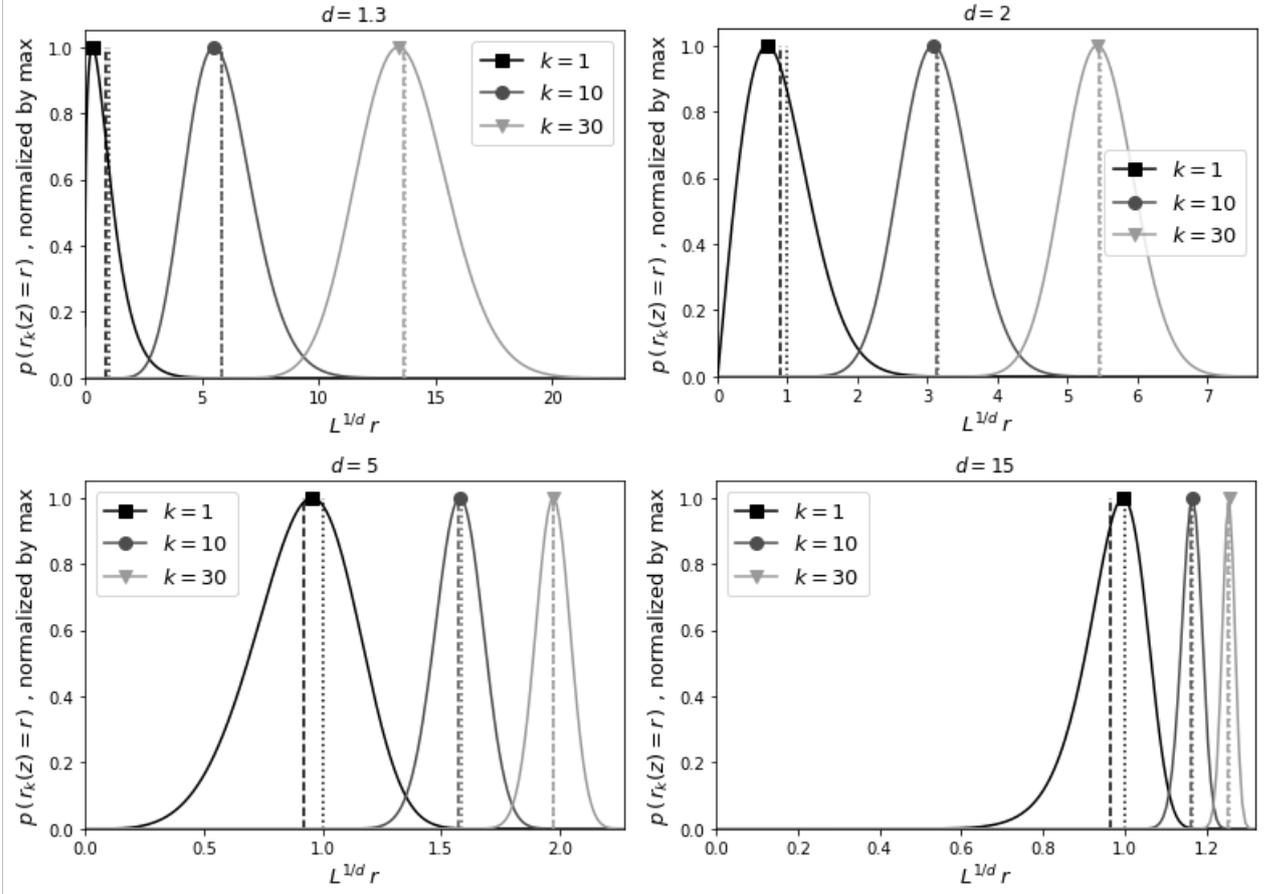
Figure II.2: Probability density functions of r_k, the k-th analog-to-target distance, for fixed values of k, and of the local dimension d, from Eq. (II.11). All densities p_k are normalized by their maximum value. From left to right and top to bottom: $d = 1.4$, 2, 5, and 10. The distances are normalized by $L^{-1/d}$. Dashed vertical lines indicate the exact mean value $\langle r_k \rangle$ from Eq. (II.12a), while dotted vertical lines indicate the approximate value $(k/L)^{1/d}$ from Eq. (II.13a). The argmax values of p_1, p_{15} and p_{30} are shown respectively with squares, circles and triangles.

2.3.5 Alternative proof for $p_k(r)$ and joint probability distribution

Lucarini et al. (2016) give a detailed analysis of the map from $\mathcal{A}$ to $\mathbb{R}$, $x \mapsto -\log \operatorname{dist}(z, x)$, using tools from dynamical systems theory and extreme value theory (EVT). For our purpose, it is interesting to look at the simpler distance map $x \mapsto \operatorname{dist}(z, x)$.

The minimum of this map over the catalog is achieved for the best analog of z, a_1. The minimum is thus r_1. EVT tells (see Coles et al., 2001) that in the limit of large catalog, the minimum of this lower-bounded distance map on a finite sample of the attractor (a catalog of size L) follows a Weibull distribution, after rescaling. The Poisson law from Eq. (II.3) with $k = 1$ actually gives the scaling and the exact form of the Weibull distribution:

$$\mathbb{P}(r_1 > r) = e^{-Lr^d},$$

for positive r, otherwise the probability is 1.

The K largest order statistics of this function then correspond to the K analogs of the point z. Again, in the limit of large catalog and for small enough K, EVT provides the limit law (see Coles et al., 2001) for the k-th minima of this distance function when $L \to \infty$:

$$\mathbb{P}(r_k > r) = e^{-Lr^d} \sum_{s=0}^{k-1} \frac{(Lr^d)^s}{s!} \, .$$

Differentiating and with a bit of rearrangement, one finds back the formula of Eq. (II.11).

$$
\begin{aligned}
p_k(r) &= -\frac{\partial}{\partial r} \mathbb{P}(r_k > r) \\
&= d\, L\, r^{d-1} \frac{\left(L\, r^d\right)^{k-1}}{(k-1)!}\, e^{-L\, r^d} \, .
\end{aligned}
$$

From a broader perspective, extremal process theory (Lamperti, 1964) gives the joint distribution of analog-to-target distances $p_{1:K}$ in the limit $L \to \infty$:

$$p_{1:K}(r_1, \ldots, r_K) = (dL)^K \left(\prod_{k=1}^{K} r_k \right)^{d-1} e^{-Lr_K^d}$$

where the function is non-zero only when $0 < r_1 < r_2 < \ldots < r_K$. For notation convenience and only in this formula, the random variables r_k are noted identically as the

values they can possibly take.

2.4 Consequences for applications of analogs

2.4.1 Comparison with previous studies

The pioneering work of Van Den Dool (1994) focuses on the minimum length of catalog needed to have a 95% chance to find at least one analog with a distance below a low threshold ε. With our notations, this condition can be written

$$L \mid \mathbb{P}(r_1 < \varepsilon) > 0.95 \,. \tag{II.15}$$

Van Den Dool (1994) uses a Gaussian approximation for the difference between two states, which is reasonable in high dimensions. Then $\mathbb{P}(r_1 < \varepsilon) = 1 - (1 - \alpha^{D_1})^L$, where α is the probability that the distance between two arbitrarily chosen states is less then ε and can be expressed as the integral of a Gaussian probability density function. For small ε, $\alpha = \mathcal{O}(\varepsilon)$ and $\alpha^{D_1} \ll 1$. This finally suggests

$$L > \frac{\log 0.05}{\log(1 - \alpha^{D_1})} \approx \frac{-\log 0.05}{\alpha^{D_1}} \tag{II.16}$$

Similar results can be found from Eq. (II.11). Indeed, one has $\mathbb{P}(r_1 < \varepsilon) = \int_0^\varepsilon p_1(r)\mathrm{d}r = 1 - \left(\exp(-\varepsilon^d)\right)^L$, so that $\alpha \approx \varepsilon$. Here, D_1 is replaced by the local finite-resolution dimension d. Thus, our analysis encompasses the one of Van Den Dool (1994).

Nicolis (1998) extended the work of Van Den Dool (1994). Interpreting Eq. (II.16) in terms of mean return times and using the formula from Kac (1959), she found an expression of mean return times using the identity $\mu_{z,r} \approx r^{D_1}$ and a mean velocity. This theoretical analysis includes neither variations in phase space of the return time, nor variability of the return time due to the variability of the catalog for fixed L. However, Nicolis (1998) performed empirical estimates of such variations of the return time, shading light on the pitfalls of an analysis limited to mean return times.

In the present paper, the point of view switches from statistics of return times to statistics of analog-to-target distance, and is extended to the K best analogs rather then just the first one. The full probability distribution of Eq. (II.11) gives a detailed view of the variability of the process of searching for analogs.

2.4.2 Searching for analogs: consequences

The full probability distribution of Eq. (II.11) has many consequences for the practical search of analogs.

For very low-dimensional systems ($D_1 < 2$), the first analog-to-target distance has a lower variability than the next ones, so that a given value of r_1 will be more representative of the next values of r_1 than a given value of r_{10} would be of the next values of r_{10}. The inverse phenomenon happens for higher dimensional systems ($D_1 > 2$). This can be taken into account to evaluate the expected performances of analog methods.

Also, the scaling $r_k \sim k^{1/d}$ implies that the growth with k of the mean analog-to-target distance is much faster for low-dimensional systems ($D_1 \lesssim 2$), so that the 30-th analog would be much farther from z than the first one. This would justify the use of a lower number of analogs K in low-dimensional spaces, while high values of K would not have a great impact on analog-to-target distances in high dimensions (see the abscissa of the lower-left panel in Fig. II.2).

For instance, Lguensat et al. (2017) use analogs to produce forecasts of several well-known dynamical systems, setting $K = 40$, while the use of Gaussian kernels with a variable bandwidth equal to $\lambda_z = \mathrm{median}_k r_k$ allows to discard analogs with $r_k > \lambda_z$. One might think that the filtering out of analogs with $r_k > \lambda_z$ make the forecast procedure relatively insensitive to the choice of K. Conversely, assuming that $\lambda_z \approx \langle r_{[K/2]} \rangle$ where $[K/2]$ is the integer part of $K/2$, we have that λ_z grows with K as $\lambda_z \sim K^{1/d}$. Thus, for low-dimensional systems such as the one of Lorenz (1963) for which $D_1 \approx 2.06$, our results suggest that high values of K would have detrimental effects on the efficiency of analog methods.

The joint distribution of analog-to-target distances from Sec. 2.3.5 theoretically allows to express the probability distributions of any random variable of the form $\sum_k \omega_k r_k^p$, where $\{\omega_k\}_k$ are weights and p is a positive integer. Such quantities can give error bounds for analog methods (see Chapter III for the case of analog forecasting). However, a closed form for the distribution of such variables is yet to be derived.

2.5 Numerical experiments

2.5.1 Three-variable Lorenz system

Using the procedure of Caby et al. (2019), one estimates the local finite-resolution dimension $d(z, r_K)$ for any point z using the K-best analogs in the system of Lorenz (1963),
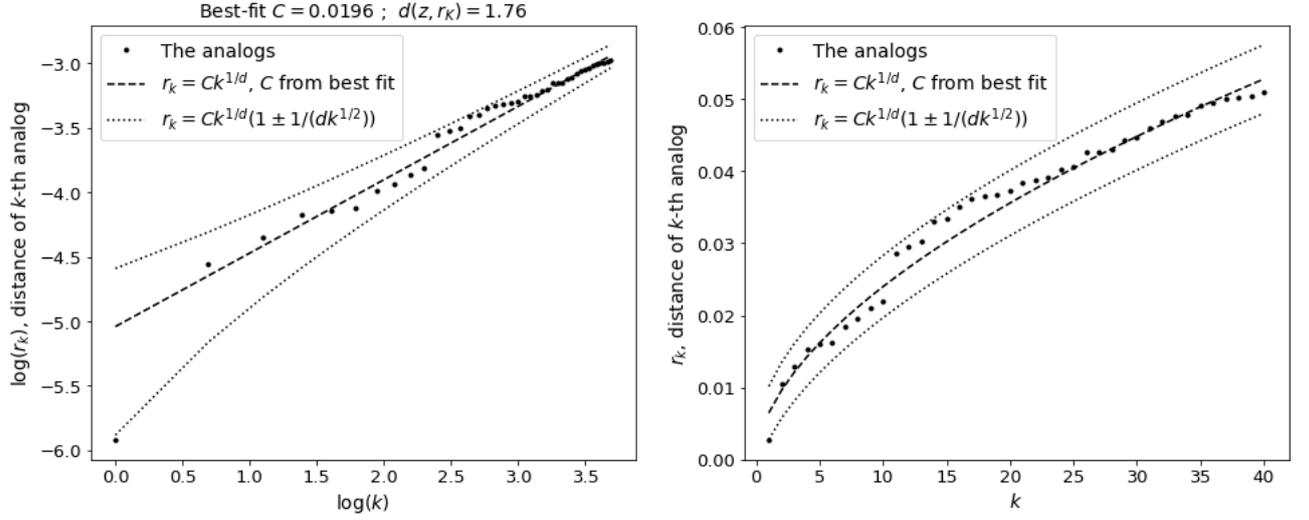
Figure II.3: Analog-to-target distance r_k, against analog number k at the same point z than in figure II.1. Full circles are the empirical points given by the analogs. The dashed dark line is the best fit from equation (II.17) where d is fixed (from Caby's method) and C is estimated with least-squares in log-scale. Assuming that this fit gives an estimation of the mean, the dotted lines represent approximate standard-deviation around this mean.

hereafter noted L63. This procedure is illustrated in Fig. II.1. Then, the scaling of Eq. (II.13a) is used to make a least-squares fit from the data

$$r_k(z) \approx^{\text{LS}} C(z)k^{1/d}, \tag{II.17}$$

where $r_k(z)$ is the observed k-th analog-to-target distance and $\approx^{\text{LS}}$ means that the constant $C(z)$ is evaluated with least-squares from Eq. (II.17). Fig. II.3 shows an application of this procedure for a given z of the L63, plotting the real values of r_k, and using $C(z)k^{1/d}$ as an approximation for $\langle r_k \rangle$ and dotted lines show the standard deviation around the mean from the approximate Eq. (II.13b).

From Eq. (II.17) and Eq. (II.13) one expects to find:

$$C(z) \approx L^{-1/d}, \tag{II.18}$$

however, as L takes large values (from 10^5 to 10^7 or more), a small estimation error for d results in a large estimation error for $L^{-1/d}$. Another way to look at this estimation issue is that d is relatively insensitive to a rescaling of distances. Let:

$$d' = \frac{\log \mu_{z,r}}{\log(r/\rho)} \tag{II.19}$$

where ρ is a scalar value and r/ρ is a rescaled version of r. Then $d' \sim d$ as long as $|\log \rho| \ll |\log r|$. In particular, the method of Caby et al. (2019) is insensitive to a rescaling, as it involves only ratios of distances (see the horizontal axis of Fig. II.1). Thus, Eq. (II.18) does not hold when C and d are determined as explained above. This is why $C(z)$ is rather evaluated through Eq. (II.17), which allows one to find the rescaling:

$$C(z) = \frac{\rho(z)}{L^{1/d}}. \tag{II.20}$$

Note that similar issues are raised by Faranda et al. (2011) regarding the continuity of $\mu_{z,r}$ with respect to r and its limiting behaviour for small r, which motivates Lucarini et al. (2014) to postulate that $\mu_{z,r}$ is the product of r^{D_1} and a slowly varying function of r, which is in some sense equivalent to our hypothesis that $C(z)$ has to be rescaled when the local dimension is estimated from the method of Caby et al. (2019).

Those formulas are tested in numerical experiments using the system of Lorenz (1963), with results reported in Fig. II.4. Analogs of a fixed target point z are sought for in 3×600 independent catalogs, with three different catalog sizes. Each catalog is built from a random draw without replacement of L points inside a (common) trajectory of 10^9 points, generated using a Runge-Kutta numerical scheme with a time step of 0.01 in usual non-dimensional notations. The dimension is calculated using $K = 150$ points, where this number is justified by a bias-variance trade-off: using this number and testing the procedure on 100 points picked from the measure μ, one finds a mean dimension between 2.03 and 2.04, which is coherent with reported values Caby et al. (2019), and a standard deviation of ~ 0.26. Using a lower value of K results in a higher variance, and using higher values results in biases that are dependent on the value of L used in this study.

The consistency of empirical densities of ρ across varying values of L validates the scaling of C with L and d. Empirical probability densities of rescaled analog-to-target distances, also consistent across varying catalog sizes, are coherent with the theoretical probability densities from Eq. (II.11). The values of the rescaling parameter ρ are not surprising, as typical values of distances between points in the attractor are ~ 16 and maximum distances are ~ 28. Note that Nicolis (1998) uses a rescaling in her study of analog return times with Lorenz's three-variable system, dividing all distances by the
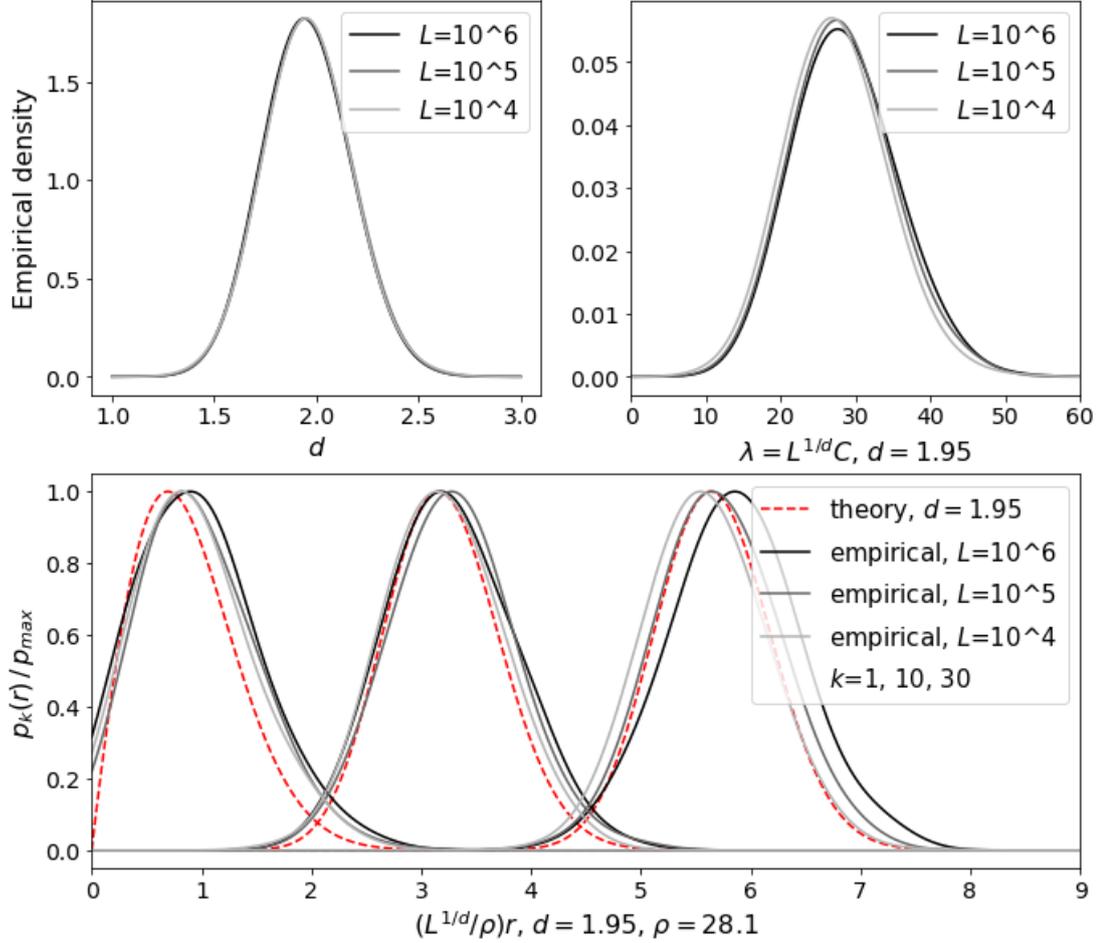
Figure II.4: Numerical experiments of the system of Lorenz (1963), for a fixed target point z, using catalogs of various sizes L, repeating the experiment 600 times for each catalog to obtain empirical probability densities. Top-left: empirical density of the local dimension d, obtained with the method of Fig II.1 and with 150 analogs. Top-right: empirical density of $\rho(z)$ obtained from Eq. (II.17) and Eq. (II.20), setting d to the mean value of its empirical densities. Bottom: normalized empirical probability densities of rescaled distances $\frac{L^{1/d}}{\rho}r$, setting ρ and d to the mean value of their empirical densities, and normalized theoretical probability densities using the same value of d. The probability densities are estimated using Gaussian kernels with bandwith of .15 (for d), 4 (for ρ), .3 (for rescaled r).

maximum distance between two points on the attractor.

Repeating this experiment for different target points z gives similar results. Values of ρ are of the same order of magnitude as the one reported in Fig. (II.4). The consistency across varying values of L is almost always recovered, except for some points that have slightly higher dimensions $d \gtrsim 2.15$ (not shown here). The authors expect this to come from a bad choice of K when estimating the dimension and the rescaling factor: the choice of $K = 150$ is relevant for most points, but should be adapted to the local dimension.

2.5.2 Wind data from Iroise sea

To further appreciate the applicability of our results to high-dimensional, real geophysical systems, the theoretical developments from Sec. 2.3 are tested on five years (2015-2019) of hourly 10m-wind output from the physical model AROME (Ducrocq et al., 2005) coupled with satellite, radar, and in-situ observations trough a variational data assimilation scheme (similar to the one of Fischer et al., 2005). The spatial domain is an evenly spaced grid above Britanny, with latitudes ranging from 47.075° to 49.3° and longitudes from -5.7° to -2.575°, and the spacing is 0.025°. To focus on wind at sea, land points are removed from the data: this last step allows for comparison with ongoing work targeted at offshore wind characterization and forecast.

From this data, one can compute local dimensions with the method of Caby et al. (2019). As the data is limited ($\sim 3 \times 10^4$ points), K is set to 40. Histograms of the local dimensions are plotted in Fig II.5. These show that the system lives in an attractor of dimension approximately between 7 and 15, with some states having local dimension likely to exceed 20. Large variations of the local dimension are not surprising, in particular Caby et al. (2019) also witness large maximum values of d when using a low number of analogs to compute local dimensions of North-Atlantic circulation data. Histograms of dimensions from winter, fall and spring are similar, while a more peaked and slightly shifted distribution is seen in summer, indicating that the system's attractor dimension is slightly higher in summer. This means that one needs a slightly larger catalog in order to find analogs of summer conditions.

An example of target state and analogs is shown in Fig. II.6. The chosen target state is a classical winter situation in Britanny, with strong eastward wind coming from the sea. Thus, good analogs are found in the catalog. It is hard to discriminate which analog is best: for such a high-dimensional system, the first analog-to-target distances are very similar.
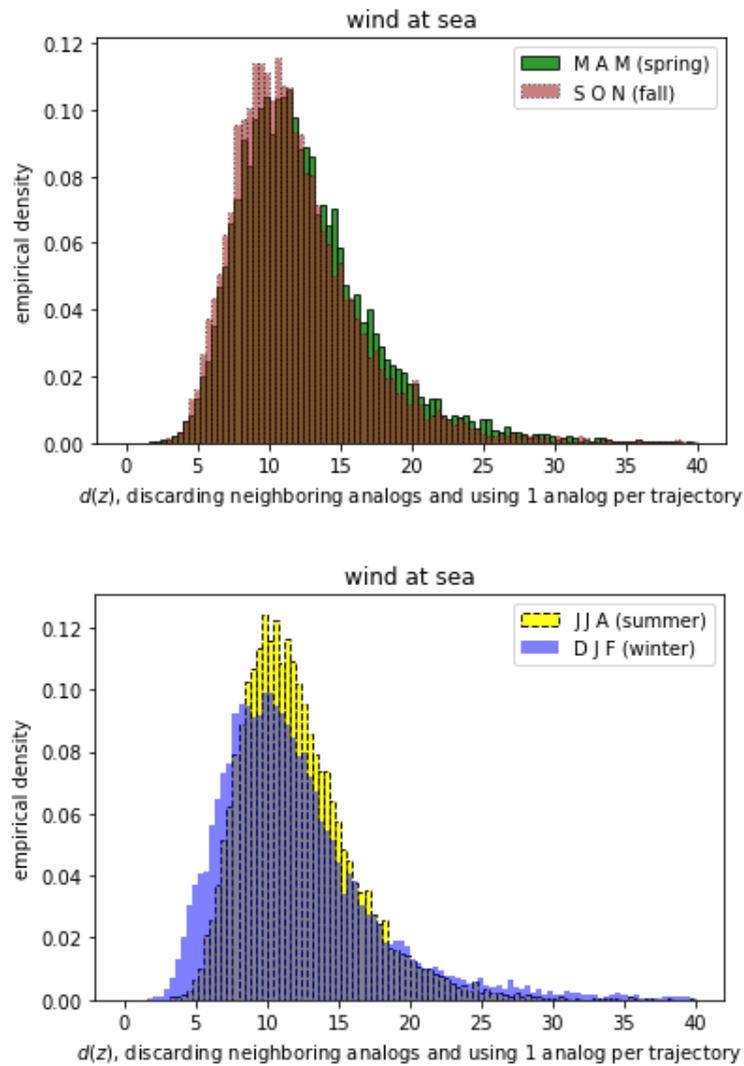
Figure II.5: Empirical probability density functions of local dimensions from 10m-wind data in the Iroise sea. We use the method of Caby et al. (2019), just like in Fig. II.1. The distributions of spring, fall (top panel) and winter (bottom panel) are similar. A small difference is seen in summer (bottom panel) with a more peaked distribution of dimensions, and a mean dimension slightly higher.
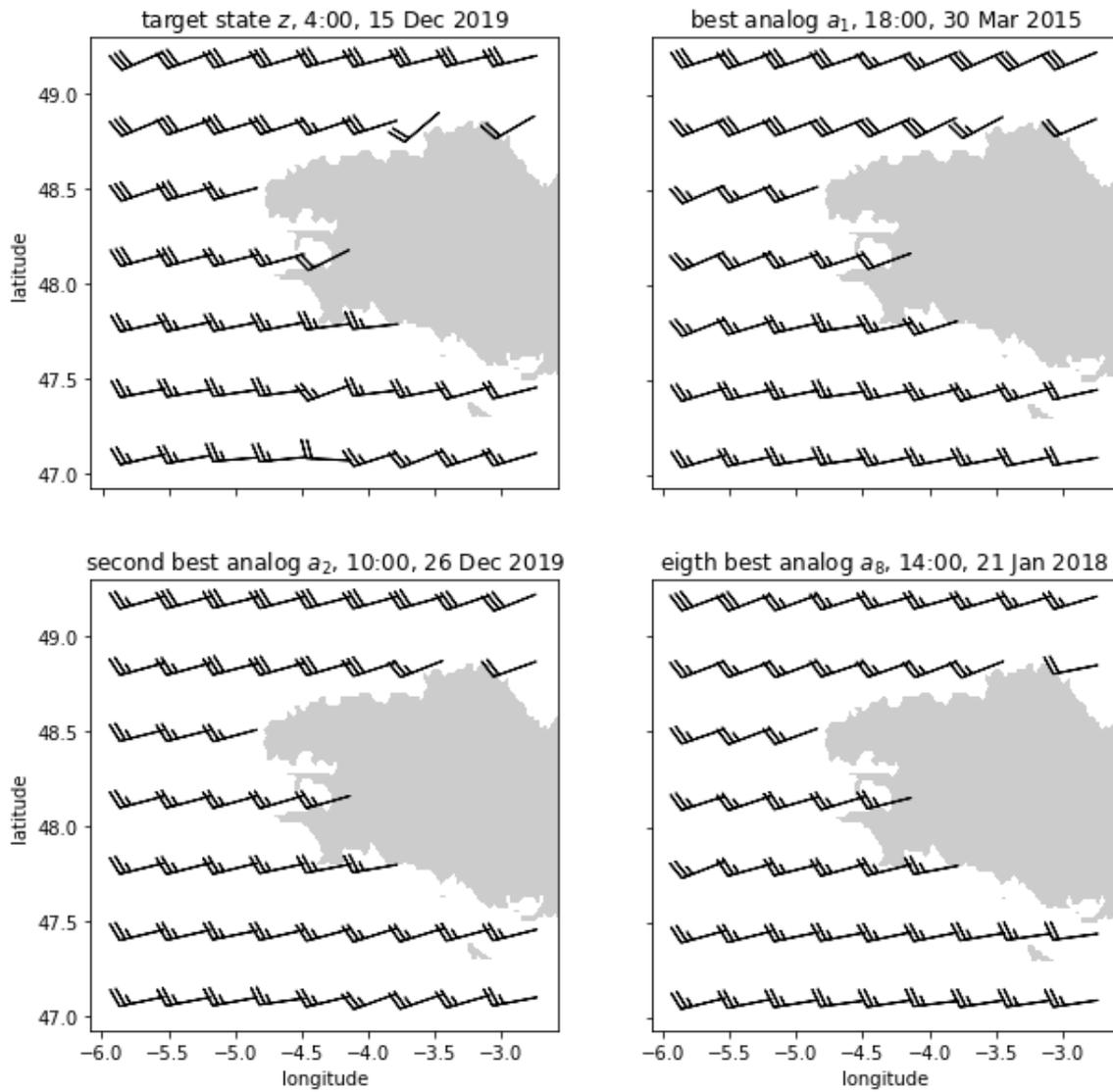
Figure II.6:   An example of target state z and first, second and eight analogs, using 10m-wind data off the coast of Britanny. Standard station model notations are used, with wind speed in knots and point-centered flags.

By rescaling the variable r_k in the following way:

$$r_k \rightarrow dk^{1/2} \left(\frac{r_k}{Ck^{1/d}} - 1 \right) , \qquad (\text{II}.21)$$

where $d(z)$ is determined through the method of Caby et al. (2019) and $C(z)$ through the least-squares approximation introduced in the previous section, one should find probability distributions approaching a standard Normal distribution, as shown in Sec. 2.3.4, especially for large values of k. However, due to the small catalog size, only probability densities up to k will be studied, otherwise the expressions obtained theoretically in the limite $L \rightarrow \infty$ are likely not to hold.

To obtain these distributions, analogs of each hourly $z \in \mathcal{C}$ ($\mathcal{C}$ is the catalog) are sought for in the catalog, omitting analogs that would be less then one and a half days away. For each z, $C(z)$ is computed from Eq. (II.17), and the distances are rescaled following Eq. II.21 and then stored. Finally, the stored values of each rescaled r_k are used to estimate probability density functions using Gaussian kernels with a bandwidth of 0.1. Fig. II.7 shows the outcome of this procedure. For comparison, a similar procedure is applied on data from the model of Lorenz (1963), using a catalog of $L = 10^6$ points and testing the procedure on 10^5 target points that are taken from a trajectory independent from the catalog. Also, the theoretical density functions from Eq. (II.14) are shown for similar (fixed) dimensions.

Note that nothing was made to remove the influence of diurnal or seasonal cycles when searching for analogs. Although the latter can have effects on the estimation of dimension (Faranda et al., 2019), this does not seem to impact the final probability distributions, as results are similar for the wind data and for the Lorenz data. Forbidding the use of analogs that are neighbors in time (if the wind at noon on the 5th of June is a good analog of z, then the wind at 1pm is likely to be a good analog as well) reduces slightly the variance of the so-found empirical density functions, and slightly shifts the empirical mean of the rescaled distances towards zero, especially for small values of k (not shown).

Fig. II.7 shows a relatively good agreement between theoretical and empirical distributions, especially for the Lorenz data. However, the empirical rescaling seems to overestimate the relative standard deviation (using $dk^{1/2}$) especially for high values of k and for the wind data. Also, the estimation for the mean using $Ck^{1/d}$ slightly biases the mean of the distributions from wind data, especially for low values of k. Variations of the parameter K did not yield better results. We also tried to remove the points for which analog-to-target distances were too high (for instance, more than 30% of the root-
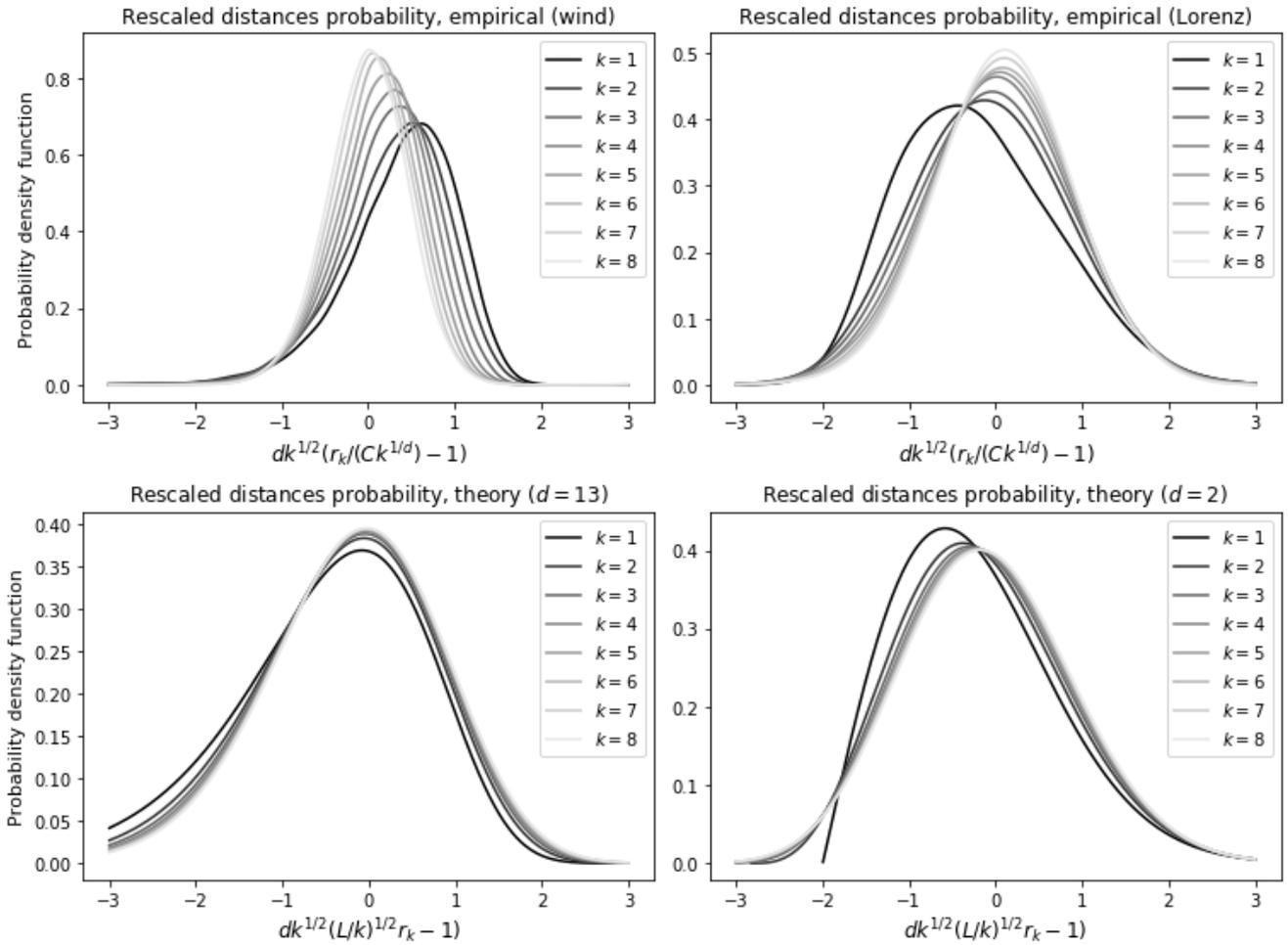
Figure II.7: Probability densities of rescaled analog-to-target distances r_k from wind data of the Iroise sea, and from numerical experiments of the Lorenz system 1963, compared to theoretical distributions from Eq. (II.14). Empirical probability densities are estimated using Gaussian kernels with a bandwidth of 0.1.

mean-squared distance between two points picked randomly in the whole dataset). This filtering procedure probably removed some data for which the assumption $L \to \infty$ was not verified, but did not change the shape of the plot significantly. With no more available data, the authors have no precise explanation for this variance estimation problem, except that it should be related to finite small catalog size which hinders precise estimation of C and d.

2.6   Conclusion

We combined extreme value theory and dynamical systems theory to derive analytical joint probability distributions of analog-to-target distances in the limit of large catalogs. Those distributions shed new light on the influence of dimension in practical use of analog. In particular, we found that the number of analogs used in empirical methods induces larger biases in low dimension than in high dimension. Contrarily to previous works on the probability to find good analogs, this study focuses on distances rather than return times, and gives whole probability distributions rather than first moments. Numerical simulations of the three-variable Lorenz system confirm the theoretical findings. 10m-wind data from the AROME physical model show that our analysis is also relevant for real systems. Our investigation indicates that these wind maps lie in an attractor of moderately high dimension ~10, such that the analog-to-target distances of the first analogs are all very similar and have a low variability.

3   Complementary analysis and perspectives

3.1   The case of partial and noisy observations

In this section the consequences of projection and additive noise are briefly examined.

Partial observations do not allow to observe the whole attractor in phase-space, but rather a projection onto a lower subspace. For instance, the wind maps used in this chapter are part of a higher-dimensional atmospheric attractor. We see no reason why the proposed methodology could not be applied to projections onto a lower subspace inside the original attractor. Numerical experiments of the system of Lorenz (1963) give results similar to Fig. II.4 after projection on the first axis (not shown here). The measured dimensions are close to 1, and theoretical curves still match using the same strategy for the estimation of dimension d and rescaling ρ. More generally, the notions of analogs and
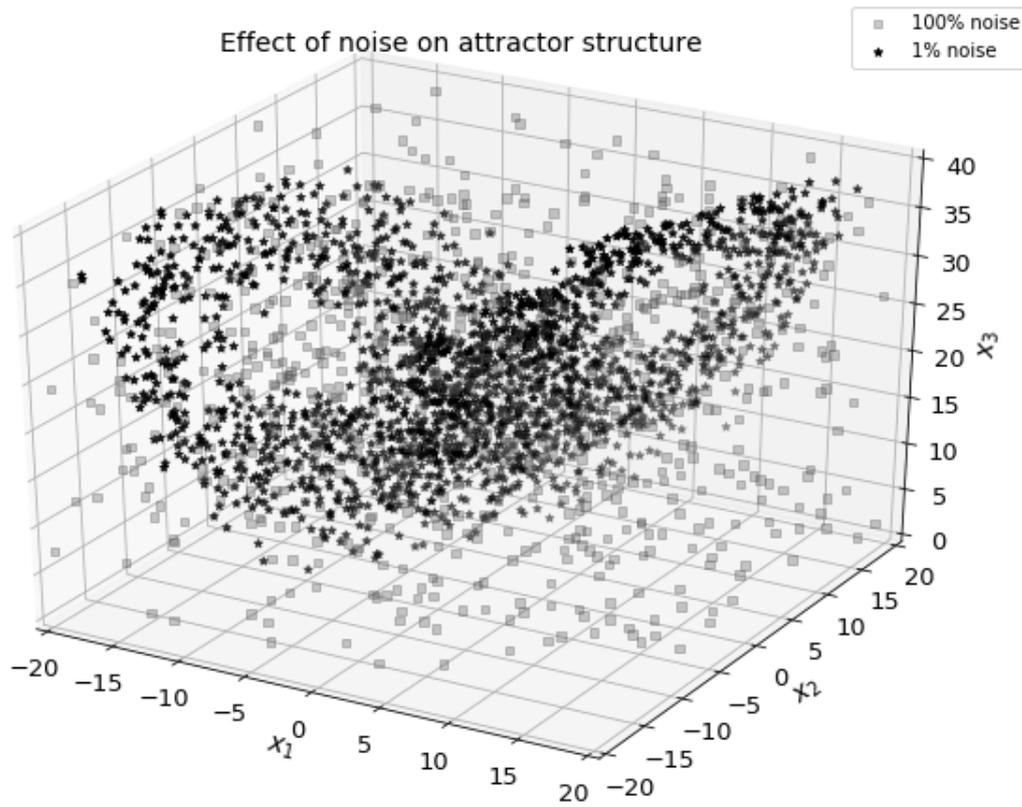
Figure II.8: The effect of isotropic additive noise on the structure of Lorenz's attractor (1963). The percentages measure the ratio of noise standard deviation and root-mean-squared distance between two points picked randomly on the attractor. Only a few thousand points are shown for readability purposes.

dimension for an observable of a dynamical system were recently studied by Caby et al. (2020).

Adding noise to observations changes the structure of the attractor. The way its structure is modified depends on the structure of the noise itself. In particular, the observed dimension is higher when adding noise in directions of phase-space that are orthogonal to the directions spanned by the attractor. Isotropic additive noise makes observations fill all directions, so that the observed dimension approaches the dimension of the observation space. To test these hypotheses, we made experiments of the system of Lorenz (1963), adding noise to the catalogs. Fig. II.8 shows the effect on the attractor's structure of various strengths of additive isotropic Gaussian noise.

Then, Fig. II.9 shows that the dimension of the noisy attractors is increased towards

3, the dimension of phase-space. Our method of estimating probability distributions for analog-to-target distances is still applicable with noisy catalogs. However, the impact of noise will be felt if one uses analogs to perform forecasting, downscaling, or for other purposes. The next chapter investigates partly the effect of noisy catalogs on analog forecasting strategies.

3.2   From analog distances to analog forecasting: an illustration

Among all applications of analogs, this PhD thesis is mostly concerned with forecasting. To illustrate the practical aspects of analog forecasting, Fig. II.10 and II.11 show the same wind maps as in Fig. II.6 but 2 hours and 12 hours later. In Fig. II.6, the target winds and the analog winds are all relatively homogeneous and coming from the west/south-west (blowing towards the east/north-east). In Fig. II.10, 2 hours later, the future state and the analogs' successors are still all quite homogeneous, with only small changes in amplitude and direction. In Fig. II.11, 12 hours later, the successors and the real future state show very different wind maps. Although the direction is still quite homogeneous in space, is varies strongly from one map to another. Also, the spatial distribution of wind amplitude can be different from a map to another (most maps show stronger winds in the north then in the south, but s_8 shows an opposite behaviour). This illustrates how analog forecasts may only be valid for small lead times, as time induces a divergence of trajectories and thus a growth of successor-to-future state distance.

For this illustrative example, many pre-processing techniques that are common in analog forecasting of atmospheric states were omitted (see for instance the algorithm of Yiou, 2014). In particular, diurnal and seasonal cycles were not forced to be respected, so that the first analog of z is a morning state while z is an evening state, and the eight analog of z is a spring state while z is a winter state. Seasonal and diurnal differences in sunlight energy forcing induce different dynamical behaviours, so that even a very good analog of z that is not from the same season or time of day can have successors very far from the actual future states of z.

However, this illustrative example shows two main aspects of analogs forecasting. First, successors diverge from the future state even when the analogs are very close to the target state. This is due to the chaotic nature of the atmosphere. Second, depending on the initial analog, the successors diverge in different directions, so that the ensemble of successors spreads out with time.

Fig. II.12 allows for further interpretation of this forecast example. The left panel

Figure II.9: Applying the aforementioned metholodogy to noisy catalogs of Lorenz's attractor (1963).
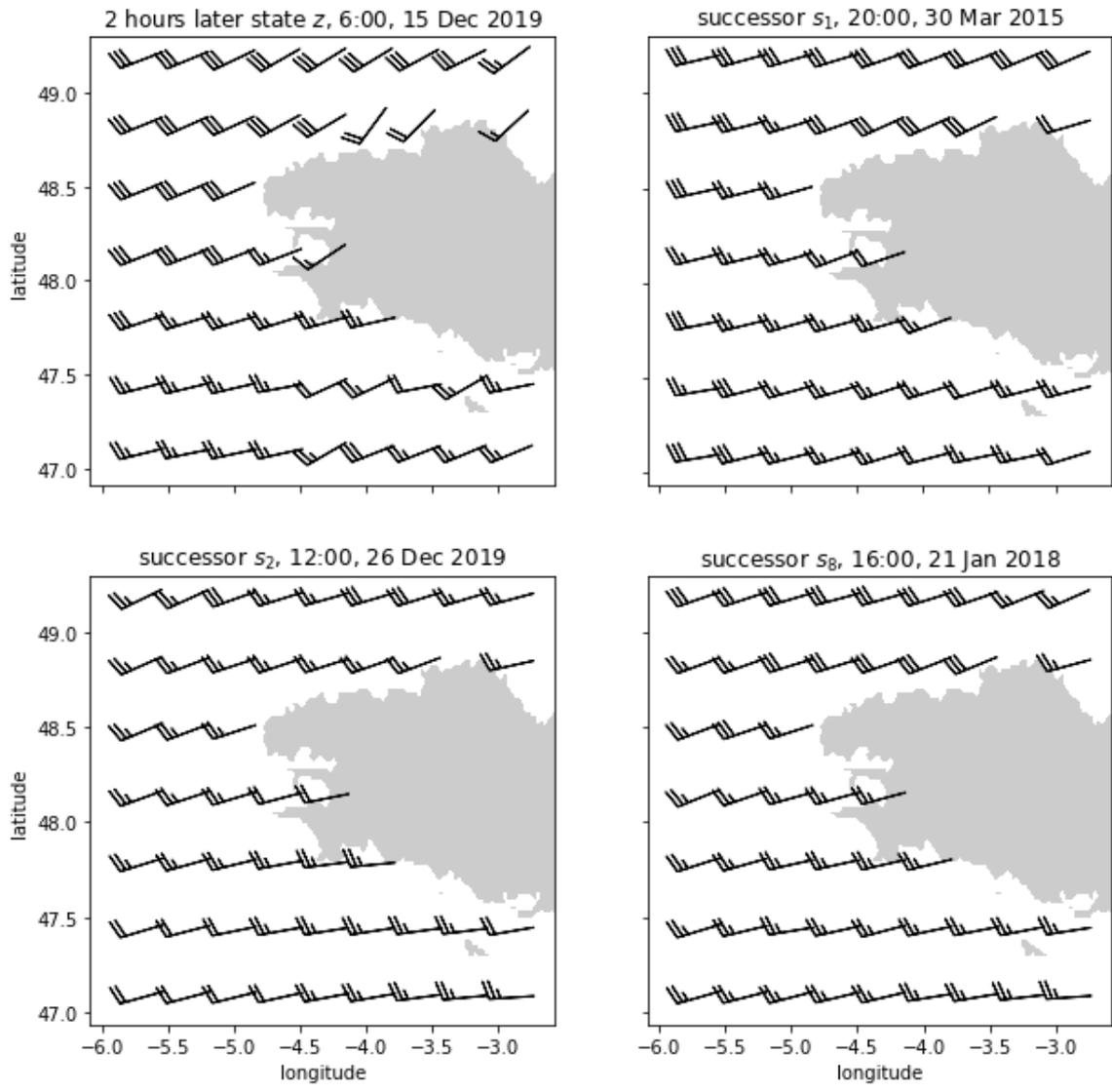
Figure II.10: Wind maps from Fig. II.6, two hours later. The target state is replaced by a "future state". Analogs are replaced by "successors" that can help estimate the "future state".
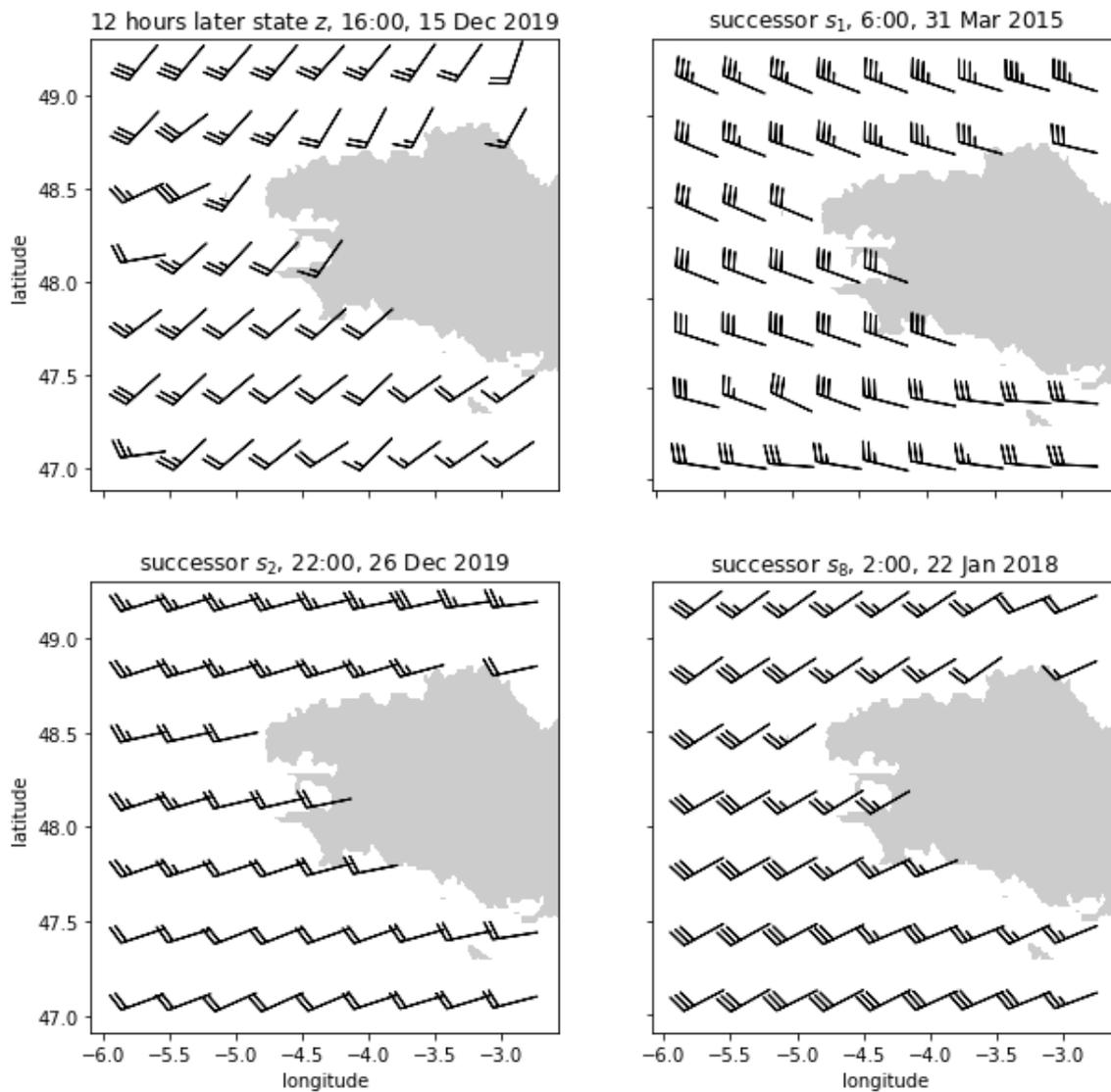
Figure II.11: Wind maps from Fig. II.6, twelve hours later. The target state is replaced by a "future state". Analogs are replaced by "successors" that can help estimate the "future state".

Example of time-divergence between successors and future state
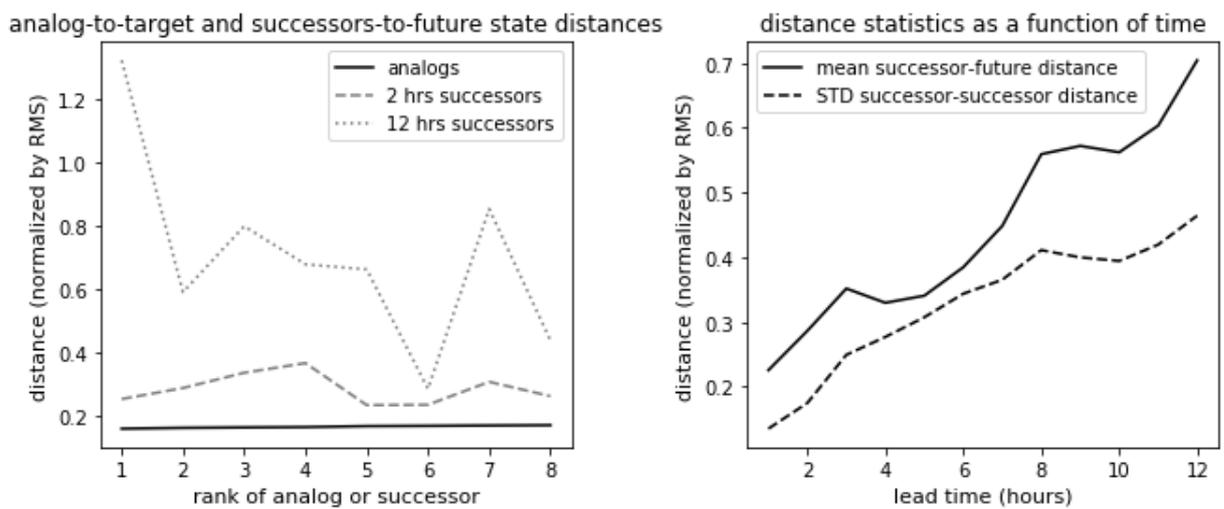
Figure II.12: All quantities plotted here correspond only to the example shown in Fig. II.6, Fig. II.10 and Fig. II.11, and not to statistics over many points. Left: distance from target wind map (respectively, future wind map) to analog wind map (respectively, successor wind map) as a function of analog rank (ordered of analog-to-target initial distances) for the lead times of two and twelve hours. Right: statistics of analog-to-target and successor-to-future state distances (full), and statistics of analog-to-analog and successor-to-successor distances (dashed).

shows that the 8 first analog-to-target distances are all below 20% of the root-mean-squared distance between two wind maps picked randomly in the whole dataset. As expected for this relatively high-dimensional dataset, the analog-to-target distances are all similar (low variance of r_k and low growth of $\langle r_k \rangle$ with k). However, the sucessors-to-future state distances show a much more complicated behaviour, even only for a forecast horizon of two hours. In particular, the successors-to-future state distance is not a growing function of k. This shows that analog-to-target Euclidean distance is not the only element that should be taken into account for forecast purposes. Other elements such as diurnal or seasonal conditions should be taken into account. Also, the dynamics gradient may not be isotropic, so that successor-to-future state distance also depends on the orientation of $a_k - z$ in feature-space. Finally, one sees that after 12 hours, the information is almost completely lost: the distance between the successors and the future state is of the order of the RMS distance between two states picked randomly in the dataset. However, some successors (the sixth one in particular) are still relatively close to the real future state, but the analog-to-target distance alone does not indicate which successor will be closest to the real future state. The right panel shows that both the mean successor-to-future state distance and the standard deviation of successor-to-successor distance are growing functions of time, at least for forecast times below 8 hours. However, it is likely that the successor-to-successor distance is slightly underestimated due to the small number of analogs considered (here, eight). On this example, the typical successor-to-successor distance is correlated with the successor-to-future state distance. Lguensat et al. (2017) use this behaviour to estimate analog forecasting errors from the covariance of successors.

To better understand these issues, the next chapter expresses successor-to-future state distances as a function of local gradients of the flow map of the associated dynamical system.

4   Summary

☐ Previous works on the chance to find analogs focused on the minimum catalog length to find one analog with distance below a given threshold. These studies revealed the exponential growth of catalog size with dimension. Here, full probability distributions of analog-to-target distances were derived, showing the influence of dimension on the second, third and next analogs.

☐ The typical scale of analog-to-target distances is proportional to the catalog size to

the power $1/d$, where d is the attractor's local dimension. The average k-th analog distance grows as $k^{1/d}$, so that the growth is faster in small dimension. The standard deviation of k-th analog-to-target distance is a growing (respectively, decreasing) function of k for $d < 2$ (respectively, $d > 2$), and a constant for $d = 2$.

☐ In applications of analog methods, the use of many analogs allows to decrease statistical variance at the cost of increased bias, due to increased analog-to-target distance. Our study reveals that this increase is faster for small dimensions, so that for medium dimensions $d \gtrsim 5$ the bias induced by using a large number of analogs (say, 50) should not be a concern.

☐ Agreement of our analytical probability distributions with numerical data from the Lorenz (1963) model is satisfactory. Empirical probability distributions from the 10m-wind AROME model output show a relatively good agreement although our analytical distributions seem to slightly overestimate the variability of analog-to-target distances.

HOW FAR FROM THE TRUTH ARE ANALOG FORECASTS?

Just as time knew to move on since the beginning
And the seasons know exactly when to change

– Stevie Wonder, *As*

1 Introducing the article

The principle behind analog forecasting is very simple: if the present state of the atmosphere resembles the state of the atmosphere on a past *analog* day (say 07:00pm, the 16th of March, 1991), then the future state of the atmosphere should also resemble the *successors* of today's analog (the weather in two hours should resemble the weather at 09:00pm, the 16th of March, 1991). If this is true, then the successors of a present state's analogs can be used to estimate the future state. Generally speaking, analog forecasting assumes that if two initial states of a dynamical system are close to each other, they will remain so for a certain amount of time. In this chapter, these assumptions are expressed in a simple mathematical framework, allowing to better interpret the behaviour of different analog forecasting strategies. We investigate the following questions:

- It is assumed that, if an analog is "close" to the initial state of a system, then its time-successors will remain "close" to the future state. Can we quantify how "close" the successors and the future state are? How is this statement dependent of the system's dynamics?

- Can we explain the behaviours of the different analog forecasting strategies proposed by Lguensat et al. (2017) based on such an analysis?

- Is it possible to predict analog forecasting errors?

To answer these questions, the following strategies and tools are used:

* The time evolution of the successors and of the real system state are described using the concept of "flow" from dynamical systems theory. In the limit of analogs being "close" to the initial state, the divergence between the successors and the future state can be described from a tangent linear model. In this framework, the successor-to-future state distance is a linear function of analog-to-initial state distance, given by the flow's Jacobian matrix.

* This allows to compare the theoretical accuracy of mean forecasts for different analog forecasting operators.

* The case of the linear regression between the analogs and the successors is examined, showing that it theoretically allows to retrieve the Jacobian matrix of the flow. This claim is investigated in numerical experiments of well-known dynamical systems.

* The link between the Jacobian matrix of the flow and the covariances of successors and analogs is expressed theoretically.

* Limitations due to additive observational noise are explored theoretically and in numerical experiments of well-known dynamical systems.

2 Article under revision for the *Journal of Atmospheric Sciences*: "Using local dynamics to explain analog forecasting of chaotic systems"

Paul Platzer[1,2,3,4] ● Pascal Yiou[1,2] ● Philippe Naveau[1,2] ● Pierre Tandeo[3] ● Yicun Zhen[3] ● Pierre Ailliot[5] ● Jean-François Filipot[4]

1. LSCE, CEA Saclay l'Orme des Merisiers, UMR 8212 CEA-CNRS-UVSQ, U Paris-Saclay, IPSL, Gif-sur-Yvette, France
2. ESTIMR - Extrèmes : Statistiques,Impacts et Régionalisation, LSCE, Gif-sur-Yvette, France
3. IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238, Plouzané, France
4. France Énergies Marines, Plouzané, France
5. Laboratoire de Mathématiques de Bretagne Atlantique, Brest, France

2.1   Abstract

Analogs are nearest neighbors of the state of a system. By using analogs and their successors in time, one is able to produce empirical forecasts. Several analog forecasting methods have been used in atmospheric applications and tested on well-known dynamical systems. Analog forecasting can be related to the dynamical equations of the system of interest. This study investigates the properties of different analog forecasting strategies by taking local approximations of the system's dynamics. We find that analog forecasting performances are highly linked to the local Jacobian matrix of the flow map, and that analog forecasting combined with linear regression allows to capture projections of this Jacobian matrix. The proposed methodology allows to estimate analog forecasting errors, and to compare different analog methods. These results are derived analytically and tested numerically on two simple chaotic dynamical systems.

2.2   Introduction

To evaluate the future state of a physical system, one strategy is to use physical knowledge to build differential equations that emulate the dynamics of this system. Then, measurements provide information on the initial state from which these equations must be integrated. Data assimilation gives a framework to account for two main types of error in this forecasting process. First, the aforementioned equations do not describe perfectly the real dynamics of the system, and solving these equations often requires additional approximations, such as numerical discretization. These first error sources refer to as model error. Second, observations are usually partial and noisy, such that the initial state from which the differential equations must be integrated is uncertain. Observation error is especially important for chaotic dynamical systems as the latter are highly sensitive to initial conditions. The quantification of model and observational uncertainties is an important topic in data assimilation (Tandeo et al., 2018).

For complex, highly nonlinear systems such as the atmosphere, forecasts based on physical equations are challenging. Indeed, the numerical integration of discretized physical equations can be greedy for high-resolution grids, and unresolved small-scale processes must be parameterized. Therefore, many empirical methods have been used in atmospheric sciences (see Van den Dool et al. 2007, and references therein). The last decades have seen a proliferation of data from numerical model outputs, observations or the combination of them (see for instance Saha et al. 2010; Hersbach et al. 2020), strengthening

scientific interest for empirical methods. One of such methods is called analog forecasting and is based on a notion originally introduced by Lorenz (1969) to estimate atmospheric predictability. Analog forecasting has been used in meteorological applications and on famous low-dimensional dynamical systems. Yiou (2014) builds an analog-based stochastic weather generator. Several authors (Tandeo et al., 2015; Hamilton et al., 2016; Lguensat et al., 2017; Grooms, 2020) combined analog forecasting and data assimilation. More generally, analog forecasting procedures are used in a large range of environmental applications, from well-known atmospheric oscillations (Alexander et al., 2017; Wang et al., 2020) to surface wind velocity (Delle Monache et al., 2013) or solar irradiance (Ayet and Tandeo, 2018).

Analog forecasting proposes to bypass physical equations and to use existing trajectories of the system instead, drawing either from numerical model output, observation data or reanalysis. Analog methods are based on the hypothesis that one is provided with many (or one long) trajectories of the system of interest. This enables to find analog states close to any initial state, and to use the time-successors of these analogs to evaluate the future state of the system. The fluctuating quality and density of available trajectories adds variability to this process. This leads to analog forecasting errors, which play the same role as the previously described model errors. The present study focuses on analog forecasting errors.

Data-based strategies have been used extensively for the forecast of dynamical systems, some of which are similar to the analog methods studied here. Early approaches similar to the locally-linear analog forecasting operator include the reconstruction of equations of motion from Crutchfield and McNamara (1987) and the forecast strategy of Farmer and Sidorowich (1987). In particular, the latter are able to estimate the scaling of forecast errors with forecast time, number of data, and intrinsic properties of the system such as Lyapunov exponents. Local approaches have also been combined with time-embeddings for data-based forecasts of dynamical systems, a strategy that has grown popular since the early works of Sauer (1994) and Sugihara (1994).

Theoretical arguments on the convergence of analog forecasts towards the real future state can be found in Zhao and Giannakis (2016), who focus on specific kernels that allow to enhance analog forecasting performances. More recently, Alexander and Giannakis (2020) showed theoretical convergence towards optimal forecasts for an extended version of conventional analog forecasting approaches, involving kernel principal component regressions. Our objective here is to relate the errors of the analog forecasting operators

of Lguensat et al. (2017) to local dynamical properties of the system. In the following, "analog forecasting" refers to the specific type of methods studied by Lguensat et al. (2017).

Preliminary results suggest that analog forecasting errors can be estimated empirically using local approximations of the true dynamics (Platzer et al., 2019). The current paper gives a more in-depth description of the theory that supports different analog forecasting procedures, and allows to evaluate the evolution of analog covariance matrices. The methodology is applied to two famous chaotic Lorenz systems. Our work has strong connection with the work of Atencia and Zawadzki (2017) who focused on the growth of analog forecasting errors to compare analogs ensembles with other ensemble forecast methods.

A theoretical framework for analog forecasting is outlined in Sec. 2.3, and three analog forecasting operators are recalled. In this framework, the successor-to-future state distance is expressed in Sec. 2.4. Finally, Sec. 2.5 examines analog forecasting mean and covariance, and investigates the link between linear regression in analog forecasting and the Jacobian matrix of the real system's flow map. The discussion section takes a broader view, outlines limitations which provide opportunities for new research. The conclusion emphasizes the major results of the paper.

2.3   Analog forecasting

2.3.1   Mathematical framework

Let a dynamical system be defined by the following time-differential equation:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}) \,, \tag{III.1}$$

where $\mathbf{x}$ is a vector that fully characterizes the state of the system, and $\mathbf{f}$ is a deterministic, vector-valued map. The space $\mathcal{P}$ in which $\mathbf{x}$ lives is called phase-space. For real geophysical systems, $\mathcal{P}$ is an infinite-dimensional function space, as variables such as pressure or temperature are continuous functions of space and time. In practice, satisfactory finite-dimensional approximations of such functions can usually be obtained through time-space discretizations. Therefore, throughout this study, $\mathcal{P}$ is a vector space of finite dimension n. The system is supposed to be autonomous, such that $\mathbf{f} : \mathcal{P} \to \mathcal{P}$ does not depend on time.

Given an initial state $\mathbf{x}_0$, a forecast gives an estimation of the state of the system $\mathbf{x}_t$

at a later time t. The true future state $\mathbf{x}_t$ is given by the flow map $\boldsymbol{\Phi} : \mathcal{P} \times \mathbb{R} \to \mathcal{P}$ such that:

$$\boldsymbol{\Phi}^t : \mathbf{x}_0 \to \boldsymbol{\Phi}^t(\mathbf{x}_0) = \mathbf{x}_t \,. \tag{III.2}$$

For the dynamical system defined through Eq. (III.1), $\boldsymbol{\Phi}$ represents the time-integration of this equation. The theorem of Poincaré (1890) states that trajectories will come back infinitely close to their initial condition after a sufficiently long time, and will do so an infinite number of time. It is valid for many geophysical systems of interest, and based on the hypothesis that the system can only access a finite volume of phase-space, and that the flow "preserves" volumes in phase-space. Furthermore, if the dynamical system has an attractor set $\mathcal{A} \in \mathcal{P}$, then almost all trajectories issued from a larger ensemble of points in phase-space (called the "basin of attraction") converge to this subset $\mathcal{A}$ (Milnor, 1985). In the following, we assume the simple case that the system has one attractor $\mathcal{A}$, and that all trajectories are in $\mathcal{A}$.

Analog methods are based on the idea that if one is provided with a long enough trajectory of the system of interest, one will find analog states close to any initial point $\mathbf{x}_0$ in the attractor $\mathcal{A}$. The trajectory from which the analogs are taken is called the "catalog" $\mathcal{C}$, and can either come from numerical model output or reprocessed observational data.

Analog forecasting thus supposes that we know a finite number of initial states that are close enough to $\mathbf{x}_0$ to be called "analogs", and that the flow map of the analogs resembles $\boldsymbol{\Phi}$. Therefore, the time-successors of the analogs should allow to estimate the real future state $\mathbf{x}_t$. In the following, the k-th analog and its successor are noted $\mathbf{a}_0^k$ and $\mathbf{a}_t^k$. Note that analog forecasting is intrinsically random as it depends on the catalog, which is one out of many possible trajectories. The variability in the catalog influences the ability of the analogs and successors to estimate the future state. This motivates the use of probabilistic analog forecasting operators $\boldsymbol{\Theta}$ such that:

$$\boldsymbol{\Theta}^t : \mathbf{x}_0 \to \boldsymbol{\Theta}^t(\mathbf{x}_0)$$

where $\boldsymbol{\Theta}^t(\mathbf{x}_0)$ is a distribution that gives information both about the estimation of the future state $\mathbf{x}_t$ and the variability of this estimation process.

Note that for chaotic dynamical systems, any small but finite perturbation in the initial condition will grow exponentially in time. Therefore, however small the distance between the analog $\mathbf{a}_0^k$ and the initial state $\mathbf{x}_0$ may be, the distance between the successor $\mathbf{a}_t^k$ and

the future state $\mathbf{x}_t$ will eventually be as large as the typical distance between two random points on the attractor. The exact divergence time is a function of both the Lyapunov time (the inverse of the largest Lyapunov exponent) and the initial distance between $\mathbf{a}_0^k$ and $\mathbf{x}_0$. After this divergence time, the analog forecast is not more informative then a simple climatological forecast (i.e., a forecast based on average statistics on the attractor). This study is devoted to the properties of analog forecasts before this exponential divergence is complete, so that $\mathbf{a}_t^k$ and $\mathbf{x}_t$ are still close to each other. Hereinafter, we says that "the exponential time-divergence is complete" when the distance between $\mathbf{a}_t^k$ and $\mathbf{x}_t$ is as large as the typical distance between any two points on the attractor.

2.3.2 Analog forecasting operators

Here are recalled three analog forecasting operators originally introduced in Lguensat et al. (2017). A finite number K of analogs $(\mathbf{a}_0^k)_{k \in [1,K]}$ and successors $(\mathbf{a}_t^k)_{k \in [1,K]}$ are used, and are assigned weights $(\omega_k)_{k \in [1,K]}$. This allows to give more weight to the pairs of analogs and successors that are best suited for the estimation of $\mathbf{x}_t$. For more details on relevant kernels, see Zhao and Giannakis (2016). The present article studies the properties of analog forecasting without restriction to any particular choice of weights and distance.

The distributions of each analog forecast $\mathbf{\Theta}^t(\mathbf{x}_0)$ is multinomial, with each pair of analog/successor defining an element of the empirical distribution.

The locally-constant operator (LC) uses only the successors to estimate $\mathbf{x}_t$.
$\mathbf{\Theta}_{\mathrm{LC}}^t(\mathbf{x}_0) \sim \sum_k \omega_k \boldsymbol{\delta}_{\mathbf{a}_t^k}(\cdot)$. The mean forecast is thus $\boldsymbol{\mu}_{\mathrm{LC}} = \sum_k \omega_k \mathbf{a}_t^k$. The covariance of the forecast is $\mathrm{cov}_{\omega_k}(\mathbf{a}_t^k)$, the ω-weighted empirical covariance of the successors.

The locally-incremental operator (LI) uses $\mathbf{x}_0$, the analogs and the successors to estimate $\mathbf{x}_t$.
$\mathbf{\Theta}_{\mathrm{LI}}^t(\mathbf{x}_0) \sim \sum_k \omega_k \boldsymbol{\delta}_{\mathbf{x}_0 + \mathbf{a}_t^k - \mathbf{a}_0^k}(\cdot)$. The mean forecast is $\boldsymbol{\mu}_{\mathrm{LI}} = \mathbf{x}_0 + \sum_k \omega_k (\mathbf{a}_t^k - \mathbf{a}_0^k)$. The covariance of the forecast is $\mathrm{cov}_{\omega_k}(\mathbf{a}_t^k - \mathbf{a}_0^k)$, the ω-weighted empirical covariance of the increments.

The locally-linear operator (LL) performs a weighted linear regression between the analogs and the successors. The regression is applied between $\mathbf{a}_0^k - \boldsymbol{\mu}_0$ and the successors $\mathbf{a}_t^k$, where $\boldsymbol{\mu}_0 = \sum_k \omega_k \mathbf{a}_0^k$. This gives slope $\mathbf{S}$, intercept $\mathbf{c}$, and residuals $\boldsymbol{\xi}^k = \mathbf{a}_t^k - \mathbf{S}\left(\mathbf{a}_0^k - \boldsymbol{\mu}_0\right) - \mathbf{c}$.
$\mathbf{\Theta}_{\mathrm{LL}}^t(\mathbf{x}_0) \sim \sum_k \omega_k \boldsymbol{\delta}_{\boldsymbol{\mu}_{\mathrm{LL}} + \boldsymbol{\xi}^k}(\cdot)$. The mean forecast is $\boldsymbol{\mu}_{\mathrm{LL}} = \mathbf{S}\left(\mathbf{x}_0 - \boldsymbol{\mu}_0\right) + \mathbf{c}$. The
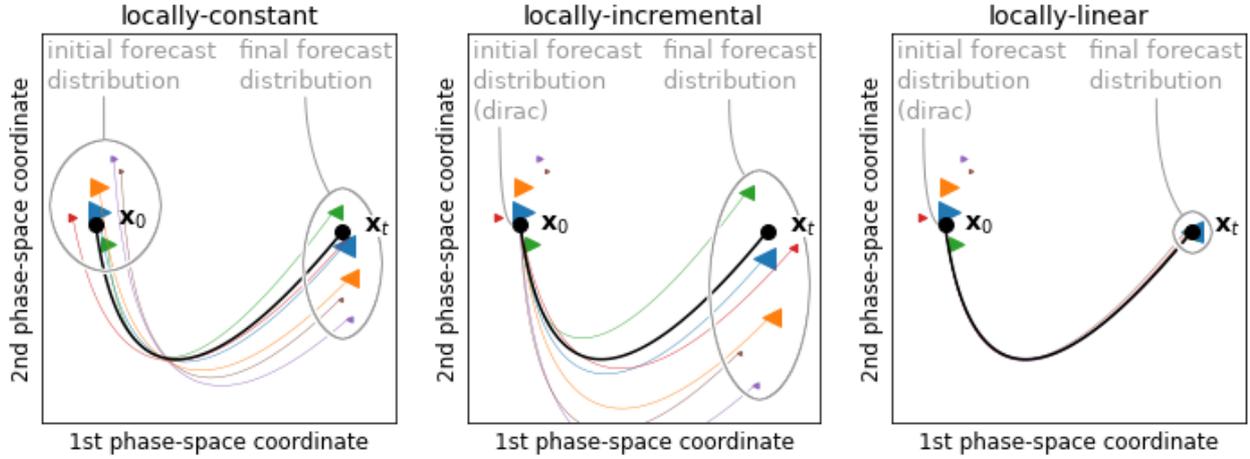
Figure III.1: Analog forecasting operators presented in Sec. 2.3.2. The flow map $\mathbf{\Phi}^t(\mathbf{x}_0)$ has a simple polynomial form. Analogs are drawn from a normal distribution centered on $\mathbf{x}_0$ and follow the same model as the real state $\mathbf{x}$. The same analogs and flow maps are used for the three operators and are represented on each panel. Weights ω_k are computed using Gaussian kernels. The real initial and future states $\mathbf{x}_0$ and $\mathbf{x}_t$ are displayed in full circles. On the left panel, analogs are in colored, right-pointing triangles, and successors in left-pointing triangles with the same colors. The size of the k-th triangle is proportional to the weight ω_k. In the middle and right panels, the elements of the forecast distribution at time t are also in colored, left-pointing triangles.

covariance of the forecast is $\mathrm{cov}_{\omega_k}(\boldsymbol{\xi}_k)$, the ω-weighted empirical covariance of the residuals.

The locally-constant (LC), locally-incremental (LI) and locally-linear (LL) analog forecasting operators are illustrated in Fig. III.1. The variance of the LC is similar around $t = 0$ and for the final value of t. On the other hand, the variance of the LI goes to 0 as $t \to 0$, but for large times the LI estimator has a larger variance compared to the LC. The next sections provide some information that help interpreting this phenomenon. The LL is able to catch the dynamics, and therefore shows a small variance and a good precision at all times. This is due to the fact that, in this example, non-linear terms are small and the flow map of the analogs matches exactly the real system's flow map.

It is worth mentioning another kind of analog forecasting operator called "constructed analogs" (CA). It is a particular case of the locally-constant operator where the weights ω_k^{CA} can have negative values and are such that the mean of the analogs $\boldsymbol{\mu}_0$ is as close as possible to the initial state: $\left\{\omega_k^{\mathrm{CA}}\right\}_k = \mathrm{argmin}_{\{\omega_k\}_k}|\sum_k \omega_k \mathbf{a}_0^k - \mathbf{x}_0|$. It was used by Van Den Dool (1994) to create better analogs in the case of small catalogs. Later, Tipett and

DelSole (2013) showed that CA are equivalent to the locally-linear operator with constant weights. In the following and unless otherwise specified, it is assumed that the weights ω_k are positive and decreasing functions of the distance between $\mathbf{a}_0^k$ and $\mathbf{x}_0$.

2.4 Successor-to-future state distance

2.4.1 Notations and hypotheses

This work assumes that the evolution dynamics of the analogs are similar to the evolution dynamics of the system of interest, and that the system is deterministic. This can be stated in a differential equation form:

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}) \\[2mm] \mathbf{x}_{t=0} = \mathbf{x}_0 \end{cases}, \tag{III.3a}$$

$$\forall k, \begin{cases} \dfrac{\mathrm{d}\mathbf{a}^k}{\mathrm{d}t} = \mathbf{f}_a(\mathbf{a}^k) \\[2mm] \mathbf{a}_{t=0}^k = \mathbf{a}_0^k \end{cases}, \quad \text{with} \quad \mathbf{f}_a = \mathbf{f} + \delta\tilde{\mathbf{f}}, \tag{III.3b}$$

or in an integrated form using flow maps:

$$\mathbf{x}_t = \mathbf{\Phi}^t(\mathbf{x}_0), \tag{III.4a}$$

$$\forall k, \ \mathbf{a}_t^k = \mathbf{\Phi}_a^t(\mathbf{a}_0^k), \quad \text{with} \quad \mathbf{\Phi}_a^t = \mathbf{\Phi}^t + \delta\tilde{\mathbf{\Phi}}^t, \tag{III.4b}$$

where $\mathbf{\Phi}_a$ is the flow map of the analogs, and $\tilde{\mathbf{\Phi}}$ is the difference between the analog and real flow maps normalized through the scalar value δ such that $\mathbf{\Phi}$, $\mathbf{\Phi}_a$ and $\tilde{\mathbf{\Phi}}$ are of the same order of magnitude. The maps $\mathbf{f}$, $\mathbf{f}_a$ and $\tilde{\mathbf{f}}$ are defined accordingly.

In these formulations, the fundamental hypotheses of analog forecasting are the continuity of $\mathbf{\Phi}^t$ (or $\mathbf{f}$) with respect to the phase-space variable $\mathbf{x}$, the density of the catalog $\mathcal{C}$ (for all k, $\mathbf{a}_0^k$ is close to $\mathbf{x}_0$ for a given metric) and the adequacy of the analogs' dynamics (δ is small, $\mathbf{\Phi}_a \approx \mathbf{\Phi}$).

The next section will investigate the ability of analogs and successors to approximate the real system state, provided that t is before the exponential time-divergence is complete, and that the aforementioned hypotheses are verified.

2.4.2   When analogs work : Taylor expansions of the dynamics

Distance between successor and real state   Assuming different levels of smoothness of the flow maps and using Taylor expansions, one can estimate the difference between the real future state $\mathbf{x}_t$ and any given successor $\mathbf{a}_t^k$ at leading order:

$$\mathbf{a}_t^k - \mathbf{x}_t = \delta\tilde{\mathbf{\Phi}}^t(\mathbf{x}_0) + \left[\mathbf{\nabla}\mathbf{\Phi}^t|_{\mathbf{x}_0}\right] \cdot (\mathbf{a}_0^k - \mathbf{x}_0) + \mathcal{O}\left(|\mathbf{a}_0^k - \mathbf{x}_0|^2, \ \delta|\mathbf{a}_0^k - \mathbf{x}_0|\right) \qquad \text{(III.5a)}$$

where $\mathbf{\nabla}\mathbf{\Phi}^t|_{\mathbf{x}_0}$ is the Jacobian matrix (the matrix of partial derivatives in phase-space) of $\mathbf{\Phi}^t$ at $\mathbf{x}_0$, '$\cdot$' is the matrix multiplication, and $\mathcal{O}\left(|\mathbf{a}_0^k - \mathbf{x}_0|^2, \ \delta|\mathbf{a}_0^k - \mathbf{x}_0|\right)$ represents higher-order terms. Neglecting these higher-order terms and lightening notations, this equation can be rewritten:

$$\mathbf{a}_t^k - \mathbf{x}_t \approx \delta\tilde{\mathbf{\Phi}}^t + \mathbf{\nabla}\mathbf{\Phi}^t \cdot (\mathbf{a}_0^k - \mathbf{x}_0),$$

where the evaluation of $\delta\tilde{\mathbf{\Phi}}^t$ and $\mathbf{\nabla}\mathbf{\Phi}^t$ at $\mathbf{x}_0$ is implicit. The leading-order difference terms explicitly described in the right hand-side of Eq. (III.5a) come from two sources. The first source is the difference between the analog and true flow maps at point $\mathbf{x}_0$, which is independent of $\mathbf{a}_0^k$. The second source of difference is the mismatch in the initial condition, left-multiplied by the Jacobian matrix of the true flow map at point $\mathbf{x}_0$.

Eq. (III.5a) states that at first order, these two error terms are additive. This is not true at higher orders. Higher-order terms include the bilinear product of $\mathbf{a}_0^k - \mathbf{x}_0$ with a matrix of second derivatives of $\mathbf{\Phi}^t$ called the Hessian, and the product of the Jacobian of $\tilde{\mathbf{\Phi}}^t$ at $\mathbf{x}_0$ and $\mathbf{a}_0^k - \mathbf{x}_0$.

Fig. III.2 shows applications of Eq. (III.5a) to the three-variable system of Lorenz (1963), hereafter noted L63. A real trajectory is compared with successor trajectories of the 10th and 100th analogs. The L63 system is solved numerically using a fourth-order Runge-Kutta finite-difference scheme, with numerical integration time step $\Delta t = 0.01$ non-dimensional time. For notation details, see Eq. (V.9) in appendix. The real trajectory has parameters $\sigma = 10$, $\rho = 28$, $\beta = 8/3$, while the σ parameter for the analog dynamics is slightly perturbed with $\sigma_a = 9 = 0.9\sigma$. The matrices $\delta\tilde{\mathbf{\Phi}}^t$ and $\mathbf{\nabla}\mathbf{\Phi}^t$ are estimated numerically using formulae given below and time step $\Delta t = 0.01$. The 10-th analog stays close enough to the real trajectory all the time (upper-left panel of Fig. III.2), therefore Eq. (III.5a) gives a satisfactory approximation of $|\mathbf{a}_t^{10} - \mathbf{x}_t|$ (upper-right panel). The

100-th analog starts to be too far from the real trajectory around $t \approx 0.7$ (upper-left and right panels), and Eq. (III.5a) provides a poor approximation of $|\mathbf{a}_t^{100} - \mathbf{x}_t|$ (upper-right panel).

The different right-hand side-terms of Eq. (III.5a) are projected on the first axis of phase-space and displayed in the lower-left panel of Fig. III.2. The "flow map" term $\delta \tilde{\boldsymbol{\Phi}}^t$ is the same for both analogs, and comes from the misspecified σ parameter. The "initial condition" term $\boldsymbol{\nabla}\boldsymbol{\Phi}^t \cdot (\mathbf{a}_0^k - \mathbf{x}_0)$ is much larger for the 100-th analog, and one can see that those terms are proportional, here negatively correlated.

Further assuming that t is small, one can express Eq. (III.5a) in the alternative formulation:

$$\boxed{\mathbf{a}_t^k - \mathbf{x}_t = t\delta\tilde{\mathbf{f}}(\mathbf{x}_0) + [\mathbf{I} + t\boldsymbol{\nabla}\mathbf{f}|_{\mathbf{x}_0}] \cdot (\mathbf{a}_0^k - \mathbf{x}_0) + \mathcal{O}\left(t^2, \; |\mathbf{a}_0^k - \mathbf{x}_0|^2, \; \delta|\mathbf{a}_0^k - \mathbf{x}_0|\right)} \, ,$$
$$\text{(III.5b)}$$

where $\mathbf{I}$ is the identity matrix. Using lighter notations, this becomes:

$$\mathbf{a}_t^k - \mathbf{x}_t \approx t\delta\tilde{\mathbf{f}} + [\mathbf{I} + t\boldsymbol{\nabla}\mathbf{f}] \cdot (\mathbf{a}_0^k - \mathbf{x}_0) \, ,$$

where the evaluation of $\delta\tilde{\mathbf{f}}$ and $\boldsymbol{\nabla}\mathbf{f}$ at $\mathbf{x}_0$ is implicit. This last formulation is analogous to a Euler scheme used in finite-difference numerical methods for solving differential equations, it is therefore valid only for small times. In the lower-right panel of Fig. III.2, one can see that the right-hand side terms of Eq. (III.5b) only approximate the terms of Eq. (III.5a) for $t \lesssim 0.1$.

Link between the two formulations, f and $\boldsymbol{\Phi}^t$ Eq. (III.5b) is a first-order expansion in time of Eq. (III.5a) . The fundamental resolvent matrix $\mathbf{M}(t, t')$ gives a more complete relationship between the two representations. $\mathbf{M}(t, t')$ is solution to the time-varying linear system $\frac{\mathrm{d}\mathbf{M}(t,t')}{\mathrm{d}t} = \boldsymbol{\nabla}\mathbf{f}|_{\mathbf{x}_t} \cdot \mathbf{M}(t, t')$ with $\mathbf{M}(t', t') = \mathbf{I}$. The fundamental resolvent matrix can be approximated numerically as $\mathbf{M}(t, t') \approx \exp(\Delta t \nabla\mathbf{f}_t) \cdot \exp(\Delta t \nabla\mathbf{f}_{t-\Delta t}) \dots \exp(\Delta t \nabla\mathbf{f}_{t'})$ with numerical time-step Δt and using the short notation $\nabla\mathbf{f}_t := \nabla\mathbf{f}|_{\mathbf{x}_t}$.

We have:

$$\delta\tilde{\boldsymbol{\Phi}}^t(\mathbf{x}_0) \approx \delta \int_0^t \mathbf{M}(t, t') \cdot \tilde{\mathbf{f}}(\mathbf{x}_{t'}) \, \mathrm{d}t' \, , \qquad \text{(III.6a)}$$
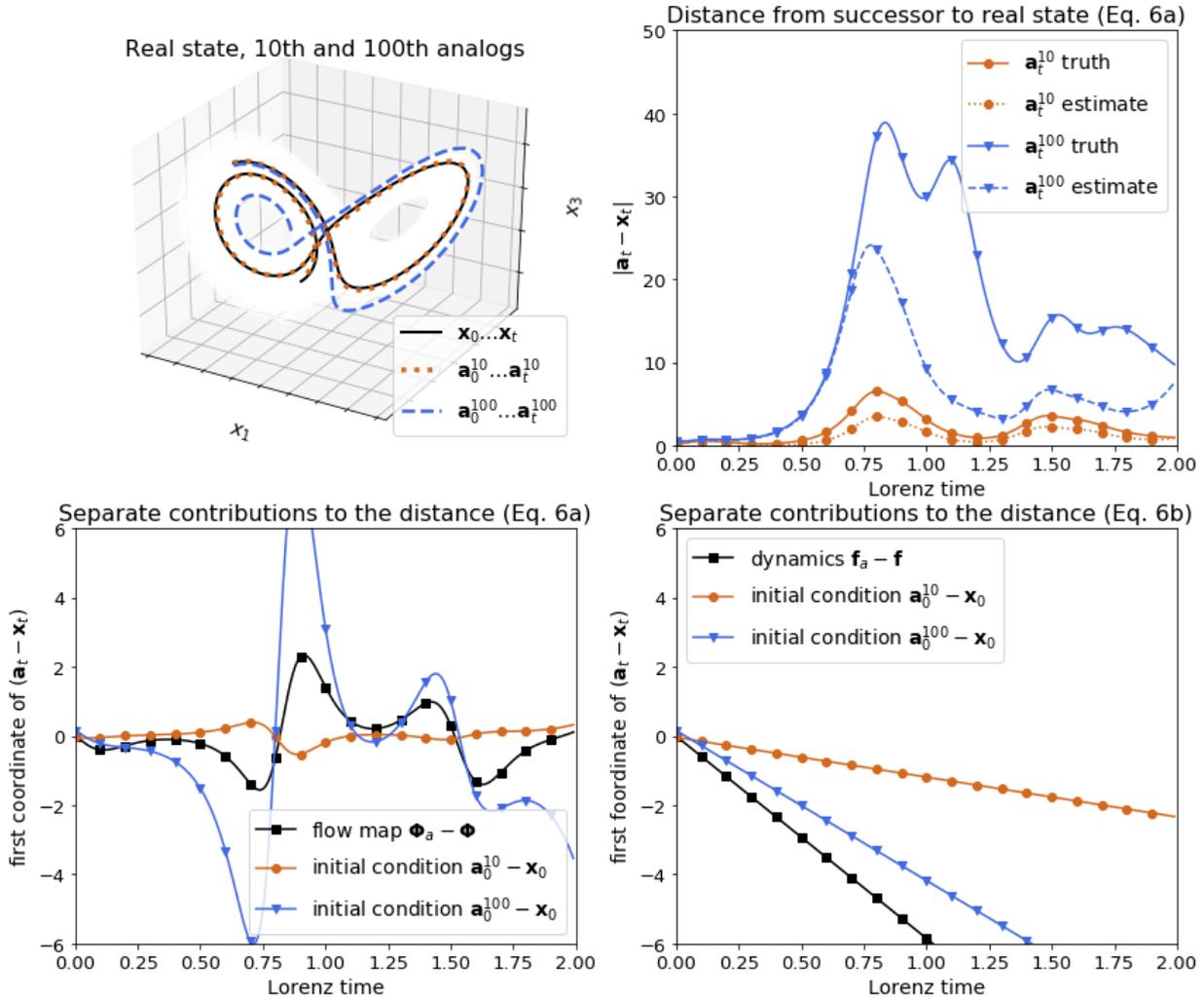
Figure III.2: Illustrating Eq. (III.5a,b) on the three-variable L63 system. Upper-left: A real trajectory from $\mathbf{x}_0$ to $\mathbf{x}_t$ and two analog trajectories, namely the 10-th best analog $\mathbf{a}_0^{10}$ to $\mathbf{a}_t^{10}$ and the 100-th best analog $\mathbf{a}_0^{100}$ to $\mathbf{a}_t^{100}$. The catalog is shown in white. Upper-right: comparing the exact value of the norm of $\mathbf{a}_t - \mathbf{x}_t$ (full lines) and the sum of the two terms on the right-hand side of Eq. (III.5a) (dashed lines). Lower-left: Contributions of the first term (black squares) and the second term (brown circles and blue triangles) of the right-hand side of equation (III.5a) projected on the first coordinate of the L63 system. Lower-right: Contributions of the first term (black squares) and the second term (brown circles and blue triangles) of the right-hand side of equation (III.5b) projected on the first coordinate of the L63 system.

90

$$\boldsymbol{\nabla\Phi}^t|_{\mathbf{x}_0} = \mathbf{M}(t, 0), \tag{III.6b}$$

where the "$\approx$" sign is here to say that Eq. (III.6a) is valid only at first order in δ. This first order is enough to compute the right-hand side terms of Eq. (III.5a), which is also valid at first order in δ.

From Eq. (III.6b) one can use Taylor developments relating $\boldsymbol{\nabla}\mathbf{f}$ and $\boldsymbol{\nabla\Phi}^t$, such as:

$$\boldsymbol{\nabla\Phi}^t = \mathbf{I} + t\,\boldsymbol{\nabla}\mathbf{f}_0 + t^2\left((\boldsymbol{\nabla}\mathbf{f}_0)^2 + \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\nabla}\mathbf{f}_0\right) + \mathcal{O}(t^3), \tag{III.7}$$

where $\boldsymbol{\nabla\Phi}^t$ is implicitly evaluated at $\mathbf{x}_0$. The short notation $\boldsymbol{\nabla}\mathbf{f}_0$ is used for $\boldsymbol{\nabla}\mathbf{f}_{\mathbf{x}_0}$, and $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\nabla}\mathbf{f}_0$ is the time derivative along the trajectory $\mathbf{x}_t$ of the Jacobian of $\mathbf{f}$, at $t = 0$. $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\nabla}\mathbf{f}|_0 := \lim_{t\to0}(\boldsymbol{\nabla}\mathbf{f}_t - \boldsymbol{\nabla}\mathbf{f}_0)/t$. At first order in t, one recovers the result expressed in Eq. (III.5b).

2.5 Consequences for analog forecasting operators

2.5.1 Mean error of analog forecasting operators

By multiplying equations (III.5a,b) by ω_k and summing over k, one can compare the distances from $\mathbf{x}_t$ to the averages $\boldsymbol{\mu}_{\mathrm{LC}}$, $\boldsymbol{\mu}_{\mathrm{LI}}$ and $\boldsymbol{\mu}_{\mathrm{LL}}$ of the different analog forecasting operators of Sec. 2.3.2. Those averages depend on t, although only implicitly in the notation. Letting $\boldsymbol{\mu}_0 = \sum_k \omega_k \mathbf{a}_0^k$ the weighted mean of the analogs, we have the following expressions.

Locally-constant mean error :

$$\boldsymbol{\mu}_{\mathrm{LC}} - \mathbf{x}_t = \delta\tilde{\boldsymbol{\Phi}}^t(\mathbf{x}_0) + \left[\boldsymbol{\nabla\Phi}^t|_{\mathbf{x}_0}\right]\cdot(\boldsymbol{\mu}_0 - \mathbf{x}_0) + \mathcal{O}\left(\sum_k \omega_k|\mathbf{a}_0^k - \mathbf{x}_0|^2,\ \delta\sum_k \omega_k|\mathbf{a}_0^k - \mathbf{x}_0|\right), \tag{III.8a}$$

$$\boldsymbol{\mu}_{\mathrm{LC}} - \mathbf{x}_t = t\delta\tilde{\mathbf{f}}(\mathbf{x}_0) + [\mathbf{I} + t\boldsymbol{\nabla}\mathbf{f}|_{\mathbf{x}_0}]\cdot(\boldsymbol{\mu}_0 - \mathbf{x}_0) + \mathcal{O}\left(t^2,\ \sum_k \omega_k|\mathbf{a}_0^k - \mathbf{x}_0|^2,\ \delta\sum_k \omega_k|\mathbf{a}_0^k - \mathbf{x}_0|\right). \tag{III.8b}$$

Locally-incremental mean error :

$$\boldsymbol{\mu}_{\mathrm{LI}} - \mathbf{x}_t = \delta\tilde{\boldsymbol{\Phi}}^t(\mathbf{x}_0) + \left[\boldsymbol{\nabla\Phi}^t|_{\mathbf{x}_0} - \mathbf{I}\right]\cdot(\boldsymbol{\mu}_0 - \mathbf{x}_0) + \mathcal{O}\left(\sum_k \omega_k|\mathbf{a}_0^k - \mathbf{x}_0|^2,\ \delta\sum_k \omega_k|\mathbf{a}_0^k - \mathbf{x}_0|\right), \tag{III.9a}$$

$$\boldsymbol{\mu}_{\mathrm{LI}} - \mathbf{x}_t = t\delta\tilde{\mathbf{f}}(\mathbf{x}_0) + [t\boldsymbol{\nabla}\mathbf{f}|_{\mathbf{x}_0}]\cdot(\boldsymbol{\mu}_0 - \mathbf{x}_0) + \mathcal{O}\left(t^2, \sum_k \omega_k|\mathbf{a}_0^k - \mathbf{x}_0|^2, \ \delta\sum_k \omega_k|\mathbf{a}_0^k - \mathbf{x}_0|\right).$$

$$\text{(III.9b)}$$

Using lighter notations with implicit evaluation at $\mathbf{x}_0$, this gives:

$$\boxed{\boldsymbol{\mu}_{\mathrm{LC}} - \mathbf{x}_t \approx \delta\tilde{\boldsymbol{\Phi}}^t + \boldsymbol{\nabla}\boldsymbol{\Phi}^t \cdot (\boldsymbol{\mu}_0 - \mathbf{x}_0) \approx t\delta\tilde{\mathbf{f}} + [\mathbf{I} + t\boldsymbol{\nabla}\mathbf{f}]\cdot(\boldsymbol{\mu}_0 - \mathbf{x}_0)} \ ,$$

$$\boxed{\boldsymbol{\mu}_{\mathrm{LI}} - \mathbf{x}_t \approx \delta\tilde{\boldsymbol{\Phi}}^t + \left[\boldsymbol{\nabla}\boldsymbol{\Phi}^t - \mathbf{I}\right]\cdot(\boldsymbol{\mu}_0 - \mathbf{x}_0) \approx t\delta\tilde{\mathbf{f}} + t\boldsymbol{\nabla}\mathbf{f}\cdot(\boldsymbol{\mu}_0 - \mathbf{x}_0)} \ .$$

The errors of the locally-constant and locally-incremental operators are both affected by the difference between the analog and real flow maps. This source of error cannot be circumvented unless provided with some information about $\delta\tilde{\boldsymbol{\Phi}}$. The other first-order error term is linear in $(\boldsymbol{\mu}_0 - \mathbf{x}_0)$, but when $t \to 0$, this term is of order t in the locally-incremental case. Thus, for small lead-times, as both $t \to 0$ and $\boldsymbol{\mu}_0 \to \mathbf{x}_0$ (i.e. dense catalog), the mean of the locally-incremental provides a better estimate of $\mathbf{x}_t$. This is why this operator is qualified by Lguensat et al. (2017) as more "physically-sound" than the locally-constant: the locally-incremental takes advantage of the fact that $\lim_{t\to 0} \boldsymbol{\Phi}^t = \mathbf{I}$, just as any finite-difference numerical scheme does. Formulas similar to Eq. (III.8-III.9) were used by Platzer et al. (2019) to predict analog forecasting errors with LC and LI operators, on the famous three-variable L63 system, with $\delta = 0$.

Another interesting property of the locally-incremental is that it can give estimates of $\mathbf{x}_t$ out of the convex hull of the catalog. This is related to what is called "novelty creation" in the machine-learning terminology. Such a property is interesting, but it also enables some inconsistent forecasts. Indeed, if t is not small enough, the locally-incremental operator can produce forecasts that have a large error due to the $-\mathbf{I}$ term in Eq. (III.9a). In Fig. III.1, one can see that the LI has a larger variance than the LC for large times.

Eq. (III.8) is also valid for constructed analogs (CA) introduced in section 2.3.2, where the weights $\{\omega_k^{\mathrm{CA}}\}$ are chosen so that $|\sum_k \omega_k^{\mathrm{CA}}\mathbf{a}_0^k - \mathbf{x}_0|$ is as small as possible. This means that the $(\boldsymbol{\mu}_0 - \mathbf{x}_0)$-linear term of equation (III.8) is also small. As mentioned earlier, Tipett and DelSole (2013) showed that this strategy is equivalent to making a linear regression. This explains why the $(\boldsymbol{\mu}_0 - \mathbf{x}_0)$-linear term is absent from Eqs. (III.10a,b).

Locally-linear mean error :

$$\boldsymbol{\mu}_{\mathrm{LL}} - \mathbf{x}_t = \delta \tilde{\boldsymbol{\Phi}}^t(\mathbf{x}_0) + \mathcal{O}\left(\sum_k \omega_k |\mathbf{a}_0^k - \mathbf{x}_0|^2 \,,\; \delta \sum_k \omega_k |\mathbf{a}_0^k - \mathbf{x}_0|\right)\,, \qquad \text{(III.10a)}$$

$$\boldsymbol{\mu}_{\mathrm{LL}} - \mathbf{x}_t = t\delta \tilde{\mathbf{f}}(\mathbf{x}_0) + \mathcal{O}\left(t^2 \,,\; \sum_k \omega_k |\mathbf{a}_0^k - \mathbf{x}_0|^2 \,,\; \delta \sum_k \omega_k |\mathbf{a}_0^k - \mathbf{x}_0|\right)\,. \qquad \text{(III.10b)}$$

Another way to understand why the $(\boldsymbol{\mu}_0 - \mathbf{x}_0)$-linear term should disappear when using the LL is to see that the LL is estimating the local Jacobian of the flow map. Indeed, the linear regression between the analogs and the successors gives an estimation of $\boldsymbol{\nabla}\boldsymbol{\Phi}^t|_{\mathbf{x}_0}$, with an estimation error that is at least of order $\mathcal{O}(|\boldsymbol{\mu}_0 - \mathbf{x}_0|, \delta)$. Sec. 2.5.2 gives a detailed argumentation to support this claim and investigates limitations. The estimation error between the linear regression matrix and the Jacobian thus adds higher-order error terms to the right-hand side of Eqs. (III.10a,b), but these are already included in the $\mathcal{O}(\sum_k \omega_k |\mathbf{a}_0^k - \mathbf{x}_0|^2 \,,\; \delta \sum_k \omega_k |\mathbf{a}_0^k - \mathbf{x}_0|)$.

We now make the explicit link between the three operators. Recall the notations of Sec. 2.3.2: the locally-linear operator finds slope $\mathbf{S}$ and intercept $\mathbf{c}$ such that for all k, $\mathbf{a}_t^k = \mathbf{S}(\mathbf{a}_0^k - \boldsymbol{\mu}_0) + \mathbf{c} + \boldsymbol{\xi}^k$ using weighted least-square estimates. This gives $\mathbf{c} = \sum_k \omega_k \mathbf{a}_t^k = \boldsymbol{\mu}_{\mathrm{LC}}$, thus we have $\boldsymbol{\mu}_{\mathrm{LL}} = \boldsymbol{\mu}_{\mathrm{LC}} + \mathbf{S}(\mathbf{x}_0 - \boldsymbol{\mu}_0)$ and the following relations hold:

$$\boxed{\boldsymbol{\mu}_{\mathrm{LC}} = \boldsymbol{\mu}_{\mathrm{LL}}|_{\mathbf{S}=\mathbf{0}}}\,,$$

$$\boxed{\boldsymbol{\mu}_{\mathrm{LI}} = \boldsymbol{\mu}_{\mathrm{LL}}|_{\mathbf{S}=\mathbf{I}}}\,,$$

such that the locally-constant and locally-incremental operator are particular cases of the locally-linear operator. We also have $\lim_{t\to 0} \mathbf{S} = \mathbf{I}$, because for all k, $\lim_{t\to 0} \mathbf{a}_t^k = \mathbf{a}_0^k$. Thus, mean forecasts of the locally-linear and locally-incremental operators are equivalent as t approaches 0:

$$\boxed{\boldsymbol{\mu}_{\mathrm{LL}} \sim_{t\to 0} \boldsymbol{\mu}_{\mathrm{LI}}}\,.$$

This analysis shows that, in terms of mean forecast error, the locally-linear operator is more precise than the locally-incremental, and the latter is more precise than the locally-constant. These findings are in agreement with the numerical experiments of Lguensat et al. (2017).

We now investigate the link between the local Jacobian of the flow $\boldsymbol{\nabla\Phi}^t|_{\mathbf{x}_0}$, and the linear regression matrix from the locally-linear operator $\mathbf{S}$.

2.5.2 Ability of analogs to estimate local Jacobians

If analogs can estimate the Jacobian of the real system, it means that analog forecasting provides a local approximation of the real dynamics, proving the relevance of analogs for short-range forecasts. Furthermore, having an estimation of the local Jacobian can be useful in some applications such as the Extend Kalman Filter, where the Jacobian allows to estimate the evolution of the forecast covariance.

Derivation of the first order error in Jacobian estimation   It is possible to find an exact expression of the first-order error term in the estimation of the local Jacobian. Let us start with the case of perfect agreement between the real and analog flow maps: $\boldsymbol{\Phi}_a = \boldsymbol{\Phi}$, or $\delta = 0$. Then, assume that in the neighborhood of $\mathbf{x}_0$ where the analogs lie, the flow $\boldsymbol{\Phi}^t(\cdot)$ can be approximated by a quadratic function in phase-space. We then have :

$$\forall k\,, \quad \mathbf{a}_t^k = \boldsymbol{\nabla\Phi}^t(\mathbf{a}_0^k - \boldsymbol{\mu}_0) + \frac{1}{2}(\mathbf{a}_0^k - \boldsymbol{\mu}_0)\boldsymbol{\nabla}^2\boldsymbol{\Phi}^t(\mathbf{a}_0^k - \boldsymbol{\mu}_0)^{\mathrm{T}} + \mathrm{Cst}\,, \qquad \text{(III.11)}$$

where "Cst" is a constant (independent of k), and the Jacobian and Hessian of $\boldsymbol{\Phi}^t$ are implicitly evaluated at $\mathbf{x}_0$ (see the appendix for notation of product of vectors and Hessian). In the next equations, the t-superscript is dropped to lighten notations. Let $\mathbf{X}$, the matrix of the analogs minus their mean, so that the k-th row of $\mathbf{X}$ is $\mathbf{a}_0^k - \boldsymbol{\mu}_0$. Similarly, let $\mathbf{Y}$ be the matrix of the successors, with the k-th row of $\mathbf{Y}$ being $\mathbf{a}_t^k$. Eq. (III.11) thus translates into $\mathbf{Y} = \mathbf{X}\boldsymbol{\nabla\Phi}^{\mathrm{T}} + \frac{1}{2}\mathbf{X}\boldsymbol{\nabla}^2\boldsymbol{\Phi}\mathbf{X}^{\mathrm{T}}$, omitting the constant.

Now let $\boldsymbol{\Omega} = \mathrm{diag}(\omega_1, \ldots, \omega_K)$, the $(K \times K)$ diagonal matrix of the weights given to each analog in the regression. Then $\mathbf{S}$ is the weighted least-squares solution of the linear regression $\mathbf{S} = (\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}\mathbf{Y}$. With a bit of rewriting, this finally gives:

$$\mathbf{S} - \boldsymbol{\nabla\Phi} = (\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}\left[\mathbf{X}\,\boldsymbol{\nabla}^2\boldsymbol{\Phi}\,\left(\frac{1}{2}\mathbf{X} + (\boldsymbol{\mu}_0 - \mathbf{x}_0)^{\mathrm{T}} \otimes \mathbf{J}_{K,1}\right)^{\mathrm{T}}\right]\,, \qquad \text{(III.12)}$$

where $\otimes$ is the Kronecker matrix product and $\mathbf{J}_{K,1}$ is the column vector with K elements all equal to 1.

Eq. (III.12) tells us that $\mathbf{S}$ is close to the Jacobian at $\mathbf{x}_0$ up to a factor that is linear in the distance between the mean of the analogs $\boldsymbol{\mu}_0$ and the analogs $\mathbf{a}_0^k$, and another factor linear in the distance between $\boldsymbol{\mu}_0$ and $\mathbf{x}_0$. These linear error term depend on the second-order phase-space derivatives of $\boldsymbol{\Phi}$ at the point $\mathbf{x}_0$ (the Hessian of $\boldsymbol{\Phi}$, noted $\boldsymbol{\nabla}^2\boldsymbol{\Phi}$).

Conducting the same derivation but relaxing the hypothesis of $\delta = 0$, one would find the same result with an added linear error term involving the Jacobian of $\tilde{\boldsymbol{\Phi}}$. This analysis allows us to say that, if the distance between the analogs and their mean is of same order as the distance between their mean and $\mathbf{x}_0$, we have:

$$\boxed{\mathbf{S} = \boldsymbol{\nabla}\boldsymbol{\Phi}^t + \mathcal{O}\left(|\boldsymbol{\mu}_0 - \mathbf{x}_0|,\, \delta\right)}\ .$$

However, the claim that the linear regression matrix $\mathbf{S}$ is able to approximate the Jacobian $\boldsymbol{\nabla}\boldsymbol{\Phi}^t$ must be tempered by several facts. To illustrate these, the regular locally-linear analog forecasting operator will now be compared with two other strategies aimed at solving dimensionality issues.

Strategies for linear regression in high dimension Dimensionality can make analog forecasting difficult, especially when using the locally linear analog forecasting operator. Here are recalled two strategies that can be used to circumvent this issue.

The first approach uses empirical orthogonal functions (EOFs, also called principal component analysis) at every forecast step. Dimension is reduced by keeping only the first n^{eof} EOFs of the set of analogs $(\mathbf{a}_0^k)_{k\in[1,K]}$, or keeping only the n^{eof} first principal components of the matrix $\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}\mathbf{X}$.

Reducing dimension using EOFs

- Find analogs $(\mathbf{a}_0^k)_{k\in[1,K]}$ of the initial state $\mathbf{x}_0$

- Compute the n EOFs of the weighted set of analogs $(\mathbf{a}_0^k)_{k\in[1,K]}$

- Keep the n^{eof} first EOFs up to 95% total variance

- Project $\mathbf{x}_0$, $(\mathbf{a}_0^k)_{k\in[1,K]}$ and $(\mathbf{a}_t^k)_{k\in[1,K]}$ on the n^{eof} first EOFs

- Perform LL analog forecasting in this projected space

The second strategy is to perform n analog forecasts, one for each coordinate of the phase-space $\mathcal{P}$, and to assume that the future of a given coordinate only depends on the

initial values of the neighboring coordinates and not on the whole initial vector $\mathbf{x}_0$. In the model of Lorenz (1996) (hereafter noted L96), Eq. (V.10) in appendix motivates the choice of keeping only the initial coordinates $\{i-2, i-1, i, i+1, i+2\}$ to estimate the i-th future coordinate. Thus we keep only $n^{\mathrm{trunc}} = 5$ initial coordinates. Thus, the LL operator performs n linear regressions with 5 coefficients at each forecast. By combining the results of those linear regressions, one finds a $n \times n$ matrix that is sparse by construction: all elements two cells away from the diagonal are equal to zero. This was introduced in Lguensat et al. (2017) as "local analogs". In the present paper this strategy will rather be termed as "coordinate-by-coordinate" analog forecasting.

Coordinate-by-coordinate forecast

- for i from 1 to n, forecast the i-th future coordinate $\mathrm{x}_{t,i}$:

    - Condition the forecast $\Theta^t_{\mathrm{LL},i}$ on a few initial coordinates around $\mathrm{x}_{0,i}$.
      $$\Theta^t_{\mathrm{LL},i}(\mathbf{x}_0) = \Theta^t_{\mathrm{LL},i}(\, \mathrm{x}_{0,i-2}\, ,\, \mathrm{x}_{0,i-1}\, ,\, \mathrm{x}_{0,i}\, ,\, \mathrm{x}_{0,i+1}\, ,\, \mathrm{x}_{0,i+2}\,)$$
    - Find analogs of the truncated initial vector $(\, \mathrm{x}_{0,i-2}\, ,\, \mathrm{x}_{0,i-1}\, ,\, \mathrm{x}_{0,i}\, ,\, \mathrm{x}_{0,i+1}\, ,\, \mathrm{x}_{0,i+2}\,)$
    - Perform LL analog forecasting $\Theta^t_{\mathrm{LL},i}$
    - Store the coefficients of the linear regression $(\, \mathrm{S}_{i,i-2}\, ,\, \mathrm{S}_{i,i-1}\, ,\, \mathrm{S}_{i,i}\, ,\, \mathrm{S}_{i,i+1}\, ,\, \mathrm{S}_{i,i+2}\,)$

- Aggregate the coefficients into the $n \times n$ matrix $\mathbf{S}$

The next section investigates limitations to the claim that the matrix $\mathbf{S}$ from the LL operator is able to approximate the Jacobian $\boldsymbol{\nabla}\boldsymbol{\Phi}^t$, and studies the impact of dimension reduction techniques on this Jacobian estimation.

Effect of the number of analogs and the phase-space dimension First, to be able to compute $\mathbf{S}$, one must have enough analogs to perform the inversion of the matrix $\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}\mathbf{X}$, where $\mathbf{X}$ is the matrix of the analogs and $\boldsymbol{\Omega}$ the diagonal matrix of the weights. This cannot be done unless K, the number of analogs used for the forecast, is superior or equal to n, the phase-space dimension. Using the EOF or coordinate-by-coordinate strategies from the previous section, one can reduce the dimension to n^{eof} or n^{trunc}, needing only to satisfy $K \geq n^{\mathrm{EOF}}$ or $K \geq n^{\mathrm{trunc}}$.

To illustrate the practical consequences of these issues, numerical simulations of the L96 system were performed with $n = 8$. The L96 is a famous chaotic dynamical system with a flexible dimension, well suited to the purpose of this study. The governing equations were solved using a fourth-order Runge-Kutta numerical scheme with an integration time

step $\Delta t = 0.05$. A catalog was built from one long trajectory (10^4 non-dimensional times) using the real equations ($\delta = 0$). Then, analog forecasting was performed at lead time 0.05, using the LL operator on 2×10^4 test points (10^3 non-dimensional times) taken from another trajectory on the attractor (independent from the catalog). Setting the number of analogs to the limiting case $K = 9$ implies that there are just enough analogs to perform the linear regression (plus one extra analog). Even though $n = 8$ is not a very large dimension, if one is provided only with 9 good analogs, one must consider dimension reduction. Regular LL analog forecasting was compared with the combination of analog forecasting with EOFs, keeping the EOFs up to 95% variance, and with the coordinate-by-coordinate analog forecasting, with $n^{\mathrm{trunc}} = 5$.

The EOF strategy ensures that the linear regression can be performed, as it projects the phase-space $\mathcal{P}$ onto the EOFs that maximize the variance in the set of analogs. Thus, the rank of the set of analogs is likely to be equal to n^{eof} in this reduced-space. However, the EOF strategy necessarily misses some of the components of the full ($n \times n$) Jacobian matrix $\boldsymbol{\nabla}\boldsymbol{\Phi}^t$, as it gives only the estimation of a ($n^{\mathrm{eof}} \times n^{\mathrm{eof}}$) matrix. The coordinate-by-coordinate method also ensures that the linear regression can be performed as long as n^{trunc} is low enough, but is also misses some of the elements of the Jacobian matrix of the flow map. Indeed, even though the coefficients of $\boldsymbol{\nabla}\mathbf{f}$ are zero two cells away from the diagonal, this is not the case of $\boldsymbol{\nabla}\boldsymbol{\Phi}^t$. Recall that, at second-order in time, $\boldsymbol{\nabla}\boldsymbol{\Phi}^t = \mathbf{I} + t\,\boldsymbol{\nabla}\mathbf{f} + t^2 \left((\boldsymbol{\nabla}\mathbf{f})^2 + \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\nabla}\mathbf{f} \right)$. Thus, some coefficients of order t^2 will not be captured by the linear regression matrix $\mathbf{S}$ using coordinate-by-coordinate analog forecasting with $n^{\mathrm{trunc}} = 5$.

The linear regression matrix $\mathbf{S}$ is then compared with $\boldsymbol{\nabla}\boldsymbol{\Phi}^t$ for the three methods. The real value of $\boldsymbol{\nabla}\boldsymbol{\Phi}^t$ is estimated with the second-order time-expansion of Eq. (III.7) that can be computed directly from the model equations (V.10). An example is shown in Fig. III.3. In this case, the regular analog forecasting misses the Jacobian with RMSE of 2.659, because the rank of the set of analogs is too low and $\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}\mathbf{X}$ is thus not invertible. Analog forecasting combined with EOFs gives a better result as it circumvents this inversion problem, with a total RMSE between $\mathbf{S}$ and $\boldsymbol{\nabla}\boldsymbol{\Phi}^t$ of 0.193. The coordinate-by-coordinate analog forecasting gives the best solution in this case, with a RMSE of 0.095. Note that many coefficients of the matrix $\mathbf{S}$ are set to zero by construction when using the coordinate-by-coordinate method.

Then, Fig. III.4 shows empirical probability density functions for the RMSE between the elements of $\mathbf{S}$ and $\boldsymbol{\nabla}\boldsymbol{\Phi}^t$ for each of the three methods. The low number of analogs

Figure III.3: Flow map Jacobian matrix estimation with the model of Lorenz (1996). Forecast lead time is $t = 0.05$ Lorenz time, catalog length is 10^4 Lorenz times, phase-space dimension is $n = 8$. $K = 9$ analogs are used for the forecast and Gaussian kernels for the weights ω_k with shape parameter λ set to the median of analog-to-state distances $|\mathbf{a}_0^k - \mathbf{x}_0|$. Upper-left: Jacobian matrix $\boldsymbol{\nabla\Phi}^t|_{\mathbf{x}_0}$. Upper-middle: linear regression matrix $\mathbf{S}$ using regular analogs. Upper-right: difference $\mathbf{S} - \boldsymbol{\nabla\Phi}^t|_{\mathbf{x}_0}$ with regular analogs, also giving the value of RMSE below the plot. Middle panels: same but the linear regression is performed in a lower-dimensional subspace spanned by the first EOFs of the set of the $K = 9$ analogs. Lower panels: same but the linear regression is performed coordinate-by-coordinate, and assuming that the coefficients are zero two cells away from the diagonal.

98
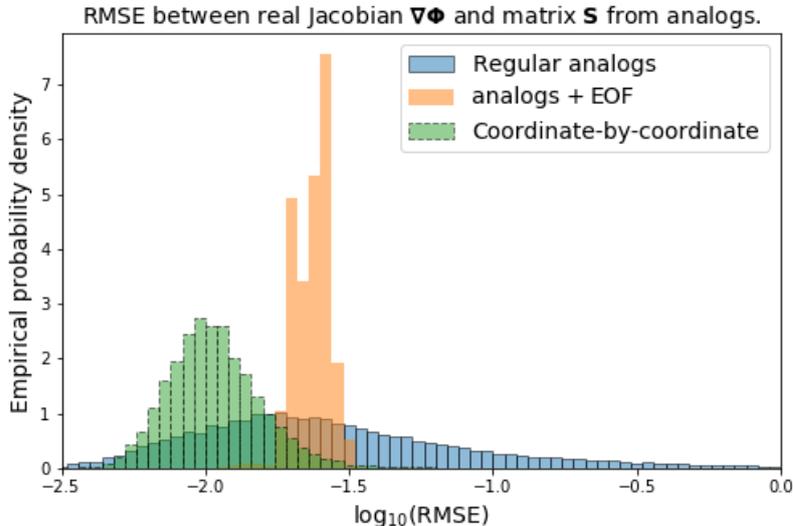
Figure III.4: Empirical probability density function of RMSE in flow map Jacobian matrix estimation, depending on the method used. We use the system of Lorenz (1996) with phase-space dimension $n = 8$. $K = 9$ analogs are used for each forecast and the methods are the same as in Fig. III.3.

implies large fluctuation of the regular LL analog forecasts, as the rank of the set of analogs used can be below or close to the phase-space dimension, making the inversion of $\mathbf{X}^\mathrm{T}\mathbf{\Omega X}$ hazardous. This variability is noticeably reduced when the inversion is performed in the n^eof-dimension reduced-space. The EOF strategy has the advantage of preventing large errors, but it does not allow very precise estimations of the Jacobian. Indeed, when using EOFs the linear regression matrix has a rank necessarily lower than n, and some information is missed. Finally, coordinate-by-coordinate analog forecasting is able to perform better estimations of the Jacobian in average, and with a variability between that of the regular analogs and that of the analogs combined with EOFs. However, the probability to have very precise estimations of the Jacobian ($\log_{10}(\mathrm{RMSE}) < -2.3$) is lower with coordinate-by-coordinate analog forecasting than with regular analog forecasting. This can be witnessed as the area under the graph for $\log_{10}(\mathrm{RMSE}) < -2.4$ is larger for regular analogs then for coordinate-by-coordinate analog forecasting. This is due to the small (order t^2) non-zero coefficients two cells away from the diagonal that the coordinate-by-coordinate analog forecasting cannot estimate.

In some situations however, the number of analogs K is much larger than the phase-space dimension n, and the linear regression matrix $\mathbf{S}$ is still unable to approximate the Jacobian $\mathbf{\nabla\Phi}^t$.

Effect of the analogs rank and the attractor's dimension   As we have seen, to calculate $\mathbf{S}$ and perform locally-linear analog forecasting, one must invert the matrix $\mathbf{X}^{\mathrm{T}}\mathbf{\Omega X}$. This means that the set of analogs must be of rank n. Yet, in some situations, the dimension of the attractor is lower than the full phase-space dimension n. Thus, if the catalog is made of one trajectory inside the attractor, the set of analogs might not be of rank n, however large K might be. In some cases, the dimension of the attractor is between $n-1$ and n, such that the matrix $\mathbf{X}^{\mathrm{T}}\mathbf{\Omega X}$ is still invertible but very sensitive to fluctuations in the rank of the set analogs.

Similar remarks can be made for the successors. If $\mathbf{Y}$ (the set of successors) is not of rank n, then the matrix $\mathbf{S}$, if it can be computed, is still not of rank n. Thus $\mathbf{S}$ will not be able to estimate the Jacobian $\boldsymbol{\nabla}\boldsymbol{\Phi}^{t}$ if the latter is of rank n. Note that the rank of the successors (the rank of the matrix $\mathbf{Y}$) is highly dependent on the rank of the analogs and the Jacobian matrix as we have $\mathbf{Y} \approx \mathbf{X}\boldsymbol{\nabla}\boldsymbol{\Phi}^{\mathrm{T}}$ at first order in $\mathbf{X}$, such that if the analogs are not of rank n, the successors are likely not to be of rank n either.

Thus, depending on the dimension of the attractor, the locally-linear analog forecasting operator might not be able to estimate the local Jacobian of the real flow map, but only a projection of this Jacobian matrix onto the local sets of analogs and successors. This is a typical case where data-driven methods are not able to reveal the full physics of an observed system unless provided with other sources of information or hypotheses, such as a parametric law.

The three-variable L63 system is used to illustrate this fact. This system is known to have a dimension of ≈ 2.06, with local variations around this value (Caby et al., 2019). This is the perfect case study where the rank of the set of analogs will be close to $n-1$. Thus, the linear regression matrix $\mathbf{S}$ between the analogs and the successors is not able to approximate the full (3×3) Jacobian matrix $\boldsymbol{\nabla}\boldsymbol{\Phi}^{t}$. Using restriction to the vector subspace spanned by the two first EOFs of the analogs $(\mathbf{e}_1^a, \mathbf{e}_2^a)$, one can understand better the connection between the two matrices $\boldsymbol{\nabla}\boldsymbol{\Phi}^{t}$ and $\mathbf{S}$. In the following, subscript "r" indicates restriction to $(\mathbf{e}_1^a, \mathbf{e}_2^a)$. The choice of using only the two first EOFs is motivated by the quasi-planar nature of the Lorenz attractor. In the next formulas the t-superscript is dropped for the sake of readability.

$$\boldsymbol{\nabla\Phi}_r = \boldsymbol{\nabla\Phi} \begin{pmatrix} \mathbf{e}_1^a \\ \mathbf{e}_2^a \\ 0 \end{pmatrix} , \qquad \mathbf{S}_r = \mathbf{S} \begin{pmatrix} \mathbf{e}_1^a \\ \mathbf{e}_2^a \\ 0 \end{pmatrix} .$$

The condition number of the set of analogs gives a direct way to measure whether the matrix $\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}\mathbf{X}$ can be inverted, and whether $\mathbf{S}$ can approximate a full rank Jacobian matrix. This number is the ratio of highest to lowest singular value. It has the advantage of being a continuous function of the set of analogs, while the rank is a discontinuous function that takes only integer values. If the condition number is large, the set of analogs is almost contained in a plane, and the analogs might not be able to approximate the full Jacobian $\boldsymbol{\Phi}$. Conversely, if the condition number is close to 1, then the rank of the set of analogs is clearly 3, and analogs will be able to approximate the full Jacobian matrix. Note that the condition number of the set of analogs is not directly linked to the dimension of the attractor. One simply uses the fact that the attractor is locally close to a plane, without referring further to the complex notion of attractor dimension.

This can be investigated through numerical simulations of the L63 system, using a fourth-order Runge-Kutta numerical scheme and a time step of $\Delta t = 0.01$ to solve the governing equations. A catalog was generated from a trajectory of 10^5 non-dimensional times, with the original equations (i.e. $\delta = 0$). Locally-linear analog forecasting was performed at horizon $t = 0.01$ with $K = 40$ analogs, on 10^4 points randomly selected on the attractor. The linear regression matrix $\mathbf{S}$ was then compared with $\boldsymbol{\nabla\Phi}^t$, with or without restriction to $(\mathbf{e}_1^a, \mathbf{e}_2^a)$. To estimate numerically the real value of $\boldsymbol{\nabla\Phi}^t$, a third-order time-expansion similar to Eq. (III.7) was computed directly from the model equations.

Fig. III.5 shows that estimation of the Jacobian by the analogs improves as the catalog size (and therefore the catalog density) grows. This validates that the analogs are able to approximate precisely the Jacobian matrix of the flow map. The figure also shows that, once restricted to the two-dimensional subspace spanned by the analogs, the estimation of the Jacobian is much more precise and less fluctuating.

Fig. III.6 displays the RMSE of the full (3×3) matrix $\mathbf{S} - \boldsymbol{\nabla\Phi}$ as a function of the condition number of the set of analogs. We can see in this figure that large RMSE values are highly correlated with high condition numbers, while low RMSE values can only be achieved when the condition number of the analogs is close to 1. Some of the

RMSE between real Jacobian $\boldsymbol{\nabla\Phi}$ and analogs matrix $\mathbf{S}$.

- (3x3) full Jacobian, median
- (2x3) Jacobian $\boldsymbol{\nabla\Phi}_r$, median
- (3x3) full Jacobian, 10%-90% qtl
- (2x3) Jacobian $\boldsymbol{\nabla\Phi}_r$, 10%-90% qtl
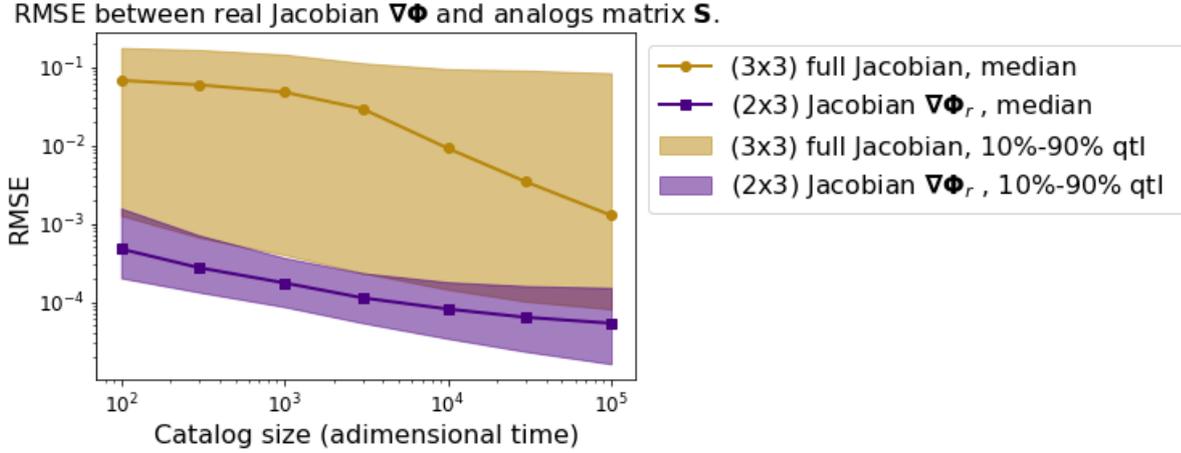
Figure III.5: RMSE in estimating with analogs the L63 Jacobian matrix, as a function of catalog size. In brown circles, the median RMSE (with 10% and 90% quantiles) of the total (3×3) Jacobian matrix. In violet squares, the median RMSE (with 10% and 90% quantiles) of the (2×2) Jacobian matrix after projection on the two first EOFs of the successors and restriction to the two first EOFs of the analogs. The number of test points decreases with catalog size, as more test points are needed for small catalogs.

remaining variability can be explained by the median distance from the analogs $\mathbf{a}_0^k$ to $\mathbf{x}_0$. Higher RMSE values are witnessed for higher median analog-to-initial state distances. This behaviour is expected: the analogs can only estimate the local Jacobian if they are close to $\mathbf{x}_0$.

All these elements show that the estimation of the Jacobian matrix from analogs is highly dependent on the number of analogs K, the condition number of the set analogs, the attractor's dimension, and the phase-space dimension n. However, the fact that the matrix from the LL operator does not approximate the full Jacobian $\boldsymbol{\nabla\Phi}^t$ does not mean that the analog forecast will poorly approximate the future state $\mathbf{x}_t$. For the LL forecast to be efficient, one only needs a good approximation of the restricted Jacobian, and that the inversion associated with the linear regression is not ill-conditioned.

2.5.3 Evolution of mean and covariance under Gaussian assumption

In this section, it is assumed that the weighted multinomial distribution of the analogs $\sum_k \omega_k \boldsymbol{\delta}_{\mathbf{a}_0^k}$ and of their successors $\sum_k \omega_k \boldsymbol{\delta}_{\mathbf{a}_t^k}$ can be approximated by Gaussian distributions:

$$\sum_k \omega_k \boldsymbol{\delta}_{\mathbf{a}_0^k} \approx \mathcal{N}\left(\boldsymbol{\mu}_0, \mathbf{P}_0\right) ,$$
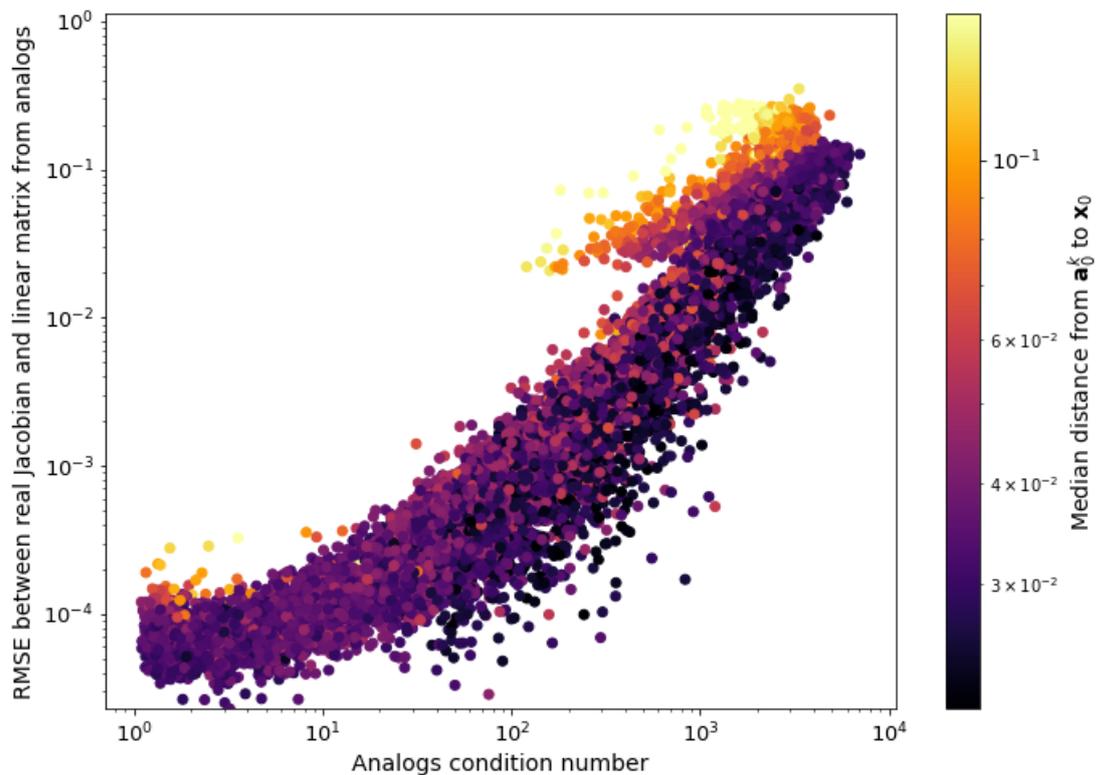
Figure III.6: RMSE in analogs estimation of the full (3x3) Jacobian matrix $\boldsymbol{\nabla\Phi}^t$ as a function of analog condition number and median analog distance, with the L63 system. The catalog size is 10^5 non-dimensional times, $\delta = 0$, and we use $K = 40$ analogs. Tests are done at 10^4 points randomly selected on the attractor.

$$\sum_k \omega_k \boldsymbol{\delta}_{\mathbf{a}_t^k} \approx \mathcal{N}\left(\boldsymbol{\mu}_t, \mathbf{P}_t\right) ,$$

where we have $\boldsymbol{\mu}_t = \boldsymbol{\mu}_{\mathrm{LC}}$. Combining these hypotheses with Eq. (III.4b) and approximating $\boldsymbol{\Phi}_a^t(\cdot)$ by its tangent around $\boldsymbol{\mu}_0$ we have the classic relationships:

$$\boldsymbol{\mu}_t = \boldsymbol{\Phi}_a^t(\boldsymbol{\mu}_0) + \mathcal{O}\left(\mathrm{Tr}\mathbf{P}_0\right) , \tag{III.13a}$$

$$\mathbf{P}_t = \boldsymbol{\nabla}\boldsymbol{\Phi}_a^t|_{\boldsymbol{\mu}_0}\,\mathbf{P}_0\,\boldsymbol{\nabla}\boldsymbol{\Phi}_a^t|_{\boldsymbol{\mu}_0}^{\mathrm{T}} + \mathcal{O}\left(\mathrm{Tr}\mathbf{P}_0\right) , \tag{III.13b}$$

where Tr is the trace operator. Similar relations can be found using the differential representation of Eq. (III.4b):

$$\frac{\mathrm{d}\boldsymbol{\mu}_t}{\mathrm{d}t} = \mathbf{f}_a(\boldsymbol{\mu}_t) + \mathcal{O}\left(\mathrm{Tr}\mathbf{P}_t\right) , \quad \boldsymbol{\mu}_{t=0} = \boldsymbol{\mu}_0 , \tag{III.14a}$$

$$\frac{\mathrm{d}\mathbf{P}_t}{\mathrm{d}t} = \boldsymbol{\nabla}\mathbf{f}_a|_{\boldsymbol{\mu}_t}\,\mathbf{P}_t + \mathbf{P}_t\,\boldsymbol{\nabla}\mathbf{f}_a|_{\boldsymbol{\mu}_t}^{\mathrm{T}} + \mathcal{O}\left(\mathrm{Tr}\mathbf{P}_t\right) , \quad \mathbf{P}_{t=0} = \mathbf{P}_0 . \tag{III.14b}$$

Now, let us make the simplifying hypothesis that $|\mathbf{x}_0 - \boldsymbol{\mu}_0|^2 \lesssim \mathrm{Tr}\mathbf{P}_0$, which means that the state $\mathbf{x}_0$ is not farther from the analogs' mean $\boldsymbol{\mu}_0$ than the standard deviation of the analogs. Then, one evaluates $\boldsymbol{\Phi}_a^t$, $\mathbf{f}_a$ and their derivatives at $\mathbf{x}_0$ and $\mathbf{x}_t$ instead of $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_t$, giving additional terms:

$$\boldsymbol{\mu}_t = \boldsymbol{\Phi}^t(\mathbf{x}_0) + \delta\tilde{\boldsymbol{\Phi}}^t(\mathbf{x}_0) + \boldsymbol{\nabla}\boldsymbol{\Phi}^t|_{\mathbf{x}_0}(\boldsymbol{\mu}_0 - \mathbf{x}_0) + \mathcal{O}\left(\mathrm{Tr}\mathbf{P}_0 , \delta|\boldsymbol{\mu}_0 - \mathbf{x}_0|\right) , \tag{III.15a}$$

$$\begin{aligned}
\mathbf{P}_t = {}& \boldsymbol{\nabla}\boldsymbol{\Phi}^t|_{\mathbf{x}_0}\,\mathbf{P}_0\,\boldsymbol{\nabla}\boldsymbol{\Phi}^t|_{\mathbf{x}_0}^{\mathrm{T}} + \delta\left(\boldsymbol{\nabla}\boldsymbol{\Phi}^t|_{\mathbf{x}_0}\,\mathbf{P}_0\,\boldsymbol{\nabla}\boldsymbol{\Phi}^t|_{\mathbf{x}_0}^{\mathrm{T}} + \boldsymbol{\nabla}\boldsymbol{\Phi}^t|_{\mathbf{x}_0}\,\mathbf{P}_0\,\boldsymbol{\nabla}\boldsymbol{\Phi}^t|_{\mathbf{x}_0}^{\mathrm{T}}\right) \\
& + \left(\left[(\boldsymbol{\mu}_0 - \mathbf{x}_0)\boldsymbol{\nabla}^2\boldsymbol{\Phi}^t|_{\mathbf{x}_0}\right]\mathbf{P}_0\boldsymbol{\nabla}\boldsymbol{\Phi}^t|_{\mathbf{x}_0}^{\mathrm{T}} + \boldsymbol{\nabla}\boldsymbol{\Phi}^t|_{\mathbf{x}_0}\mathbf{P}_0\left[(\boldsymbol{\mu}_0 - \mathbf{x}_0)\boldsymbol{\nabla}^2\boldsymbol{\Phi}^t|_{\mathbf{x}_0}\right]^{\mathrm{T}}\right) + \mathcal{O}\left(\mathrm{Tr}\mathbf{P}_0 , \delta|\boldsymbol{\mu}_0 - \mathbf{x}_0|\right) ,
\end{aligned} \tag{III.15b}$$

where terms of order $|\boldsymbol{\mu}_0 - \mathbf{x}_0|^2$ are included in $\mathcal{O}\left(\mathrm{Tr}\mathbf{P}_0\right)$ and $\boldsymbol{\nabla}^2\boldsymbol{\Phi}^t|_{\mathbf{x}_0}$ is the Hessian of $\boldsymbol{\Phi}^t$ at $\mathbf{x}_0$. In the time-differential representation we have:

$$\frac{\mathrm{d}(\boldsymbol{\mu}_t - \mathbf{x}_t)}{\mathrm{d}t} = \boldsymbol{\nabla}\mathbf{f}|_{\mathbf{x}_t}(\boldsymbol{\mu}_t - \mathbf{x}_t) + \delta\tilde{\mathbf{f}}(\mathbf{x}_t) + \mathcal{O}\left(\mathrm{Tr}\mathbf{P}_t , \delta|\boldsymbol{\mu}_t - \mathbf{x}_t|\right) , \tag{III.16a}$$

$$\frac{d\mathbf{P}_t}{dt} = \boldsymbol{\nabla}\mathbf{f}|_{\mathbf{x}_t}\,\mathbf{P}_t + \mathbf{P}_t\,\boldsymbol{\nabla}\mathbf{f}|_{\mathbf{x}_t}^{\mathrm{T}} + + \delta\left(\boldsymbol{\nabla}\tilde{\mathbf{f}}|_{\mathbf{x}_t}\,\mathbf{P}_t \,+\, \mathbf{P}_t\boldsymbol{\nabla}\tilde{\mathbf{f}}|_{\mathbf{x}_t}^{\mathrm{T}}\right)$$
$$+ \left(\left[(\boldsymbol{\mu}_t - \mathbf{x}_t)\boldsymbol{\nabla}^2\mathbf{f}|_{\mathbf{x}_t}\right]\mathbf{P}_t + \mathbf{P}_t\left[(\boldsymbol{\mu}_t - \mathbf{x}_t)\boldsymbol{\nabla}^2\mathbf{f}|_{\mathbf{x}_t}\right]^{\mathrm{T}}\right) + \mathcal{O}\left(\mathrm{Tr}\mathbf{P}_t\,,\,\delta|\boldsymbol{\mu}_t - \mathbf{x}_t|\right)\,. \quad \text{(III.16b)}$$

Eq. (III.16a) is equivalent to Eq. (III.15a), which is also equivalent to Eq. (III.5a). Eq. (III.16a) can be Taylor-expanded around $t = 0$ to find Eq. (III.8b). This analysis recovers the results from Sec. 2.5.1 for the mean forecast of the locally-constant analog forecasting operator.

Eq. (III.16b) and Eq. (III.15b) are two representations of the same phenomenon. They show that at first order, the growth in covariance between the analogs and successors is directly linked to the Jacobian matrix of $\boldsymbol{\Phi}^t$ at $\mathbf{x}_0$. The covariance of the analog forecast depends on the covariance of the analogs at $t = 0$, $\mathbf{P}_0$, and on the system's local Jacobian $\boldsymbol{\nabla}\boldsymbol{\Phi}^t|_{\mathbf{x}_0}$. This is another way to see that the analogs are highly linked to the local dynamics of the system. If the local dynamics induce a large spread in the future possible trajectories, it is captured in the successors' covariance $\mathbf{P}_t$. On the contrary, if the local dynamics are flat ($\boldsymbol{\nabla}\boldsymbol{\Phi}^t|_{\mathbf{x}_0} \simeq \mathbf{I}$ or $\boldsymbol{\nabla}\mathbf{f}|_{\mathbf{x}_0} \simeq \mathbf{0}$) the successors' covariance is equal to the analogs' covariance.

2.6   Discussion

This paper contributes to the interpretation of analog forecasting methods. Following a similar objective but using different methodology, Zhao and Giannakis (2016) set a mathematical framework for the convergence of analog forecasting operators to the flow map of the real system, with a particular emphasis on the kernels used for the weights ω_k.

There are many natural extensions to the work presented here. The first one is non-deterministic dynamics that can happen, for instance, when forecast is not performed in phase-space but in a lower-dimensional space. One might be provided only with observations of a few variables of the whole system, and try to forecast those same variables. The use of time-embeddings from Takens (1981) combined with analog forecasting is promising (Alexander et al., 2017). Also, Chau et al. (2020) build a catalog of state-space trajectories from a catalog of partial and noisy observations, using analog forecasting and data assimilation. More recently, Alexander and Giannakis (2020) proposed to combine conventional analog forecasting with projections onto a function space of observables, allowing to optimally capture the evolution of a response variable under partially observed

non-linear dynamics. The authors show the convergence of this methodology to optimal predictions in the limit of large data, using an operator-theoretic framework.

The second natural extension is to account for observation error in the catalog of analogs. As the flow map is assumed to be quasi-linear in phase-space in the neighborhood of the analogs, one could conduct the same analysis including centered additive noise for each analog and successor of the catalog, and find results similar to the ones outlined here.

One must bear in mind that the use of analog forecasting in applications implies issues such as the choice of the space in which forecasting is performed, the choice of the right metric to compare analogs and initial state, and the combination of analogs with other techniques. In data assimilation, one might want to convert the multinomial distributions of Sec. 2.3.2 to Gaussian distributions to use Kalman filtering. Ridge and Lasso regularizations could be used to ease the linear regression instead of the techniques mentioned in Sec. 2.5.2. These operational choices must be made accounting for memory use and computational time (see Lguensat et al. (2017) for differences between regular and coordinate-by-coordinate analog forecasting).

2.7   Conclusion

Analog forecasting allows to avoid solving complex nonlinear equations by using existing solutions starting from similar initial conditions. The accuracy of analog forecasting depends on local dynamical properties of the system of interest. In particular, the quality of analog forecasts is related to the Jacobian matrix of the real system's flow map, and the linear regression from analogs to successors is shown to provide an approximation of this matrix. This allows to examine the mean accuracy of known analog forecasting operators, and to compare different methods that evaluate this Jacobian matrix, using numerical experiments of famous dynamical systems. The locally-linear operator is found to give the best approximation of the future state, provided that the linear regression is not ill-posed. The locally-incremental operator is shown to give more precise forecasts at small lead times. The Jacobian matrix of the flow map is found to drive the growth of the successors' covariance matrix. Altogether, this brings theoretical evidence that analogs can be used to emulate a real system, and gives quantitative expressions for the precision of analog forecasting techniques.
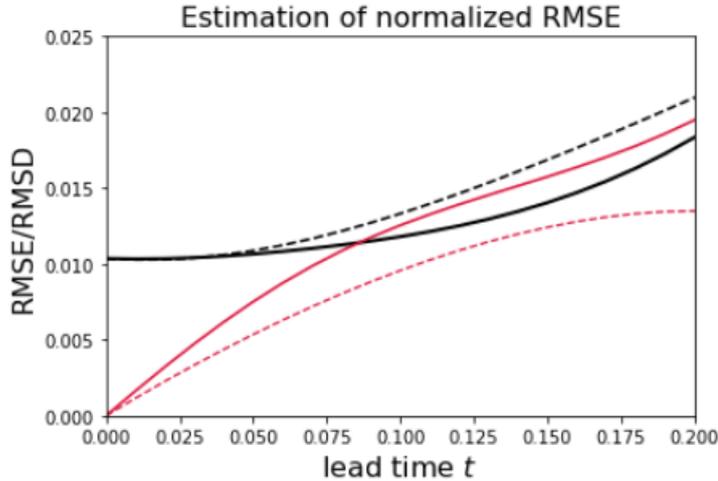
Figure III.7: RMSE associated with locally constant and locally incremental analog forecasting, normalized by the root mean squared distance between two points randomly selected on the attractor. Black, full: LC, empirical. Black, dashed: LC, predicted. Red, full: LI, empirical. Red, dashed: LI, predicted.

3   Complementary analysis

3.1   Supplementary material from conference paper

In this section are exposed supplementary results that were the topic of a peer-reviewed conference paper (Platzer et al., 2019).

In this paper, the performances of the mean locally-constant and locally-incremental operators were studied through Eq. (III.8a) and Eq. (III.9a) with $\delta = 0$, omitting the higher-order terms. The Jacobian was estimated using the second-order approximation of Eq. (III.7). Using these formulas, one can relate the RMSE of the mean analog forecasts with the root-mean-squared analog-to-target distance. This allows for an a priori estimation of the amplitude of forecast errors. Tests were performed on the system of Lorenz (1963), as reported in Fig. III.7. These show that, on average, the estimations are in good agreement with the real forecast errors.

It was then tried to combine LC and LI forecasting operators with model-based information, assuming the local Jacobian was known from the L63 equations. This greatly raised the forecast accuracy. This strategy is equivalent to using the locally-linear operator, replacing the estimated matrix **S** by the real Jacobian of the flow.

3.2 Splitted analog forecasts and finite-differences

In this section, we study the splitting of long-term analog forecasts into multiple short-term analog forecasts. To do so, we draw a parallel between analog forecasting and finite-difference numerical schemes. Application to the Lorenz (1963) system is given at the end of the section.

Finite-difference schemes numerically approach the solutions of partial differential equations that cannot be solved explicitly. They involve a discretization of continuous coordinates (time, space) and approximate the derivatives of functions along those discretized coordinates.

Let **FDS** a given finite-difference scheme that allows to compute $\tilde{\mathbf{x}}_m$, an approximation of the exact solution $\mathbf{x}(t_m)$ of Eq. (III.1) at time $t_m = m\Delta t$, through the following formula:

$$\tilde{\mathbf{x}}_{m+1} = \tilde{\mathbf{x}}_m + \Delta t \, \mathbf{FDS}(\tilde{\mathbf{x}}_m, \Delta t) \, .$$

An example of scheme is the explicit forward Euler one: $\mathbf{FDS}^{\mathrm{Euler}}(\tilde{\mathbf{x}}_m, \Delta t) = \tilde{\mathbf{f}}(\tilde{\mathbf{x}}_m)$. If no space-discretization is performed, $\tilde{\mathbf{f}} = \mathbf{f}$. Otherwise $\tilde{\mathbf{f}}$ is an approximation of $\mathbf{f}$, discretized in space. Then, three criteria are important for finite-difference numerical schemes: consistency, stability, and convergence.

Consistency (finite-differences)

The scheme must tend towards the real differential as the numerical step tends to zero. The truncation error is defined as

$$\mathbf{e}_m = \mathbf{x}(t_{m+1}) - \mathbf{x}(t_m) - \Delta t \, \mathbf{FDS}(\mathbf{x}(t_m), \Delta t) \, ,$$

with p, the order of the numerical scheme, such that $\mathbf{e}_m = \mathcal{O}(\Delta t^p)$.

Stability (finite-differences)

Applying the numerical scheme many times must not lead to divergence. In particular, fixing a time interval T over which a solution of Eq. (III.1) is sought for, one must be able to reduce the time step Δt as much as needed. For stable schemes, there exists a constant C such that for every Δt divisor of T:

$$\max_{m=0,1,\dots,T/\Delta t} \|\mathbf{FDS}^m(\,\cdot\,,\Delta t)\|_\infty < C\,,$$

Convergence (finite-differences)

The whole finite-difference approximation must tend to the exact solution of Eq. (III.1), such that for every integer $N > 0$:

$$\lim_{\Delta t \to 0} \max_{m=0,1,\dots,N} \|\mathbf{x}(t_m) - \tilde{\mathbf{x}}_m\| = 0\,.$$

This last criterion encompasses the stability and consistency criteria. The theorem of Lax and Richtmyer (1956) shows that if a finite-difference scheme is stable and consistent, then it is also convergent.

Let us now make the parallel with analog forecasting. The work of this chapter shows that mean forecasts of analog forecasting operators approach the solution of Eq. (III.1) when using large catalogs, and when analogs and real state follow the same dynamics (i.e. $\delta = 0$). However, the rate of this convergence depends on the chosen analog forecasting operator, on the lead time, and on the system's dynamics. Improved performances of the locally-incremental and locally-linear operator at small lead times motivates the splitting into multiple, small lead-time forecasts, in order to approach a long solution of Eq. (III.1) using analogs. The resulting time-discretized, approached solution will be noted $\hat{\mathbf{x}}_m$ (the hat is justified by assuming that the mean of forecasting operators is taken):

$$\hat{\mathbf{x}}_{m+1} = \hat{\boldsymbol{\Theta}}^{\Delta t}(\hat{\mathbf{x}}_m, L)\,,$$

where the dependency on the catalog size L was made explicit. Indeed, as finite-difference schemes approach the exact solution when $\Delta t \to 0$, analog forecasting operators approach the exact solution when $L \to \infty$. For simplicity, δ is assumed to be zero.

We will now define notions of consistency, order, stability and convergence for this type of multiple, mean analog forecasts.

Consistency (analog forecasting)

The application of the mean analog forecasting operator must tend towards the real future state, Δt times ahead, in the limit of large catalog. The truncation error is defined as

$$\hat{\mathbf{e}}_m = \mathbf{x}(t_{m+1}) - \hat{\boldsymbol{\Theta}}^{\Delta t}(\mathbf{x}(t_m), L).$$

where the order p of the mean analog forecasting operator is defined by $\hat{\mathbf{e}}_m = \mathcal{O}(L^{-p})$.

Sec. 2.5.1 of this chapter showed that successor-to-future state distance is at most linear in analog-to-target state distance (before Δt is too large and the exponential time divergence is complete). It was shown in Sec. 2.3 of the previous chapter, that analog-to-target state distance is of the order of $\mathcal{O}(L^{-1/d})$, where d s the local dimension, assuming that $K^{1/d}$ is not too large. It therefore follows that, in theory, mean analog forecasting operators are of order $1/d$ and $\boxed{\hat{\mathbf{e}}_m = \mathcal{O}(L^{-1/d})}$. This order is less than one, and can be relatively small even for low-dimensional systems of say, $d \approx 8$. This is consistent with the idea that very large catalogs are needed to have good analog performances in high-dimensions. However, using the locally-incremental operator makes that $\hat{\mathbf{e}}_m$ is also of order Δt, giving $\boxed{\hat{\mathbf{e}}_m = \mathcal{O}(\Delta t\, L^{-1/d})}$. Furthermore, using the locally-linear operator gives an approximation that is both quadratic in analog-to-target distance and in time. In this case the mean locally-linear analog forecasting operator is of order $2/d$ and we have $\boxed{\hat{\mathbf{e}}_m = \mathcal{O}(\Delta t^2\, L^{-2/d})}$. This proves that choosing efficient analog-forecasting methods and splitting the forecast into multiple small-time-step forecasts can greatly raise the efficiency of analog-based approximations of differential equations. Nonetheless, this would be done at the expense of added computational time.

Stability (analog forecasting)

Applying the mean analog forecasting operator many times must not lead to divergence of the numerical approximation of the exact solution. Thus, for stable mean analog forecasting operators, there exists a constant C such that:

$$\max_{m>0} \| \left(\hat{\boldsymbol{\Theta}}^{\Delta t} \right)^m (\, \cdot \, , L) \|_\infty < C \, ,$$

where $\left(\hat{\boldsymbol{\Theta}}^{\Delta t} \right)^m (\, \cdot \, , L)$ is the m-times composition of $\hat{\boldsymbol{\Theta}}^{\Delta t} (\, \cdot \, , L)$ with itself. Note that it is different from $\hat{\boldsymbol{\Theta}}^{m\Delta t} (\, \cdot \, , L)$ which corresponds to applying only one forecast at lead time $m\Delta t$.

The locally-constant operator verifies this stability criterion, as mean analog forecasts μ_{LC} are always comprised in the convex hull of the catalog. However, the locally-incremental and locally-linear could be unstable. Further investigations would have to be taken in this direction.

It is important to notice that, in practice, analog forecasting may be used in a data assimilation scheme, such that the forecast would be regularly drawn back to the real state through the added information of observations, preventing divergence. In such a context, the analog forecast would be used only for the time separating two observations.

Convergence (analog forecasting)

The maximum of the difference between the real and approximate solutions must tend towards zero in the limit of large catalogs, such that for every integer $N > 0$:

$$\lim_{L \to \infty} \max_{m=0,1,\ldots,N} \| \mathbf{x}(t_m) - \hat{\mathbf{x}}_m \| = 0 \, .$$

Perhaps, a theorem comparable to the one of Lax and Richtmyer (1956) could be established. Again, this demands further research.

Simple tests of applying $\left(\hat{\boldsymbol{\Theta}}^{\Delta t} \right)^m$ on the system of Lorenz (1963) were tried. Fig. III.8 shows the results of this small experiment for lead times up to 15 non-dimensional times. The performances of direct (no splitting into multiple forecasts) locally-constant
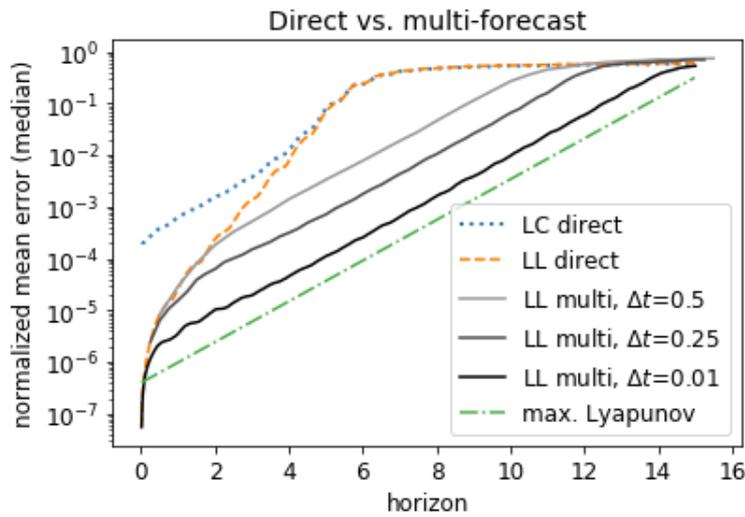
Figure III.8: Comparison between multiple forecasts $\left(\hat{\Theta}^{\Delta t}\right)^m$ and direct forecast $\hat{\Theta}^t$ with the locally-linear analog forecasting operator, on the system of Lorenz (1963), with $K{=}20$ analogs, a catalog size of $L = 10^7$ (10^5 non-dimensional times) and a time step $\mathrm{d}t = 0.01$. The direct locally-constant forecast and the slope corresponding to the maximal Lyapunov exponent are also shown. The median of errors $|\left(\hat{\Theta}^{\Delta t}\right)^m (\mathbf{x}_0) - \mathbf{x}_t|$ are computed over 10 000 points $\mathbf{x}_0$ taken randomly on the attractor.

and locally-linear operators are very different for small lead times but similar for large lead times. Splitting the forecast in multiple analog forecasts remarkably improves the performances of the mean locally-linear operator. For both direct- and multi- locally-linear operator, errors increase exponentially for large lead-times, but with a lower slope when using multiple forecasts. Using a smaller time step Δt between the analog forecasts allows to delay the beginning of the exponential error growth, such that the forecast error reaches a plateau later in time.

The exponential divergence rate of multiple locally-linear forecasts and of direct locally-constant forecasts is consistent with the inverse of the maximal Lyapunov exponent (taken here as 0.9057, see Viswanath 1998). This indicates that the error at large lead times is driven by the chaotic growth of initial truncation errors. Indeed, at each forecast step, truncation errors are only added to previous errors (linear growth in time), while the chaotic divergence induces multiplicative errors (exponential growth in time). Thus, there is a characteristic time-scale at which the chaotic errors are more important then the truncation errors. Note that finite-difference approximations of chaotic dynamical systems suffer from the same exponential divergence at large lead times.

This example shows that the notions of consistency and order are interesting for the interpretation of multiple, splitted analog forecasts. Further theoretical developments and numerical experiments would help applying the notions of stability and convergence to multiple, splitted analog forecasts.

3.3 The influence of additive noise

3.3.1 Influence on mean forecasts and Jacobian estimation

In this section, we use an error-in-variables model to evaluate the impact of additive observational noise on mean analog forecast errors and Jacobian estimation. Theoretical arguments are supported by numerical experiments of the Lorenz (1963) model.

As mentioned in the conclusion of this chapter, observations are in general both partial and noisy. Here, we first tackle the issue of noisy observations.

For this preliminary investigation, it is assumed that every element of the catalog is an observation of a state following the dynamical Eq. (III.3b) or (III.4b) with an added Gaussian noise. It is assumed that the Gaussian noises are centered, independent and identically distributed. The Gaussian noises are noted $\boldsymbol{\epsilon}_0^k$ for the k-th analog and $\boldsymbol{\epsilon}_t^k$ for the k-th successor at time t. No noise is added to the initial state $\mathbf{x}_0$ because this is taken

into account separately when treating observational noise in data assimilation schemes. From these hypothesis, Eq. (III.4) are modified to give:

$$\mathbf{x}_t = \mathbf{\Phi}^t(\mathbf{x}_0) \,, \tag{III.17a}$$

$$\forall k, \, \mathbf{a}_t^k - \boldsymbol{\epsilon}_t^k = \mathbf{\Phi}_a^t(\mathbf{a}_0^k - \boldsymbol{\epsilon}_0^k) \,, \quad \text{with} \quad \mathbf{\Phi}_a^t = \mathbf{\Phi}^t + \delta \tilde{\mathbf{\Phi}}^t \,, \tag{III.17b}$$

and Eq. (III.5a) is changed to:

$$\boxed{\mathbf{a}_t^k - \mathbf{x}_t \approx \boldsymbol{\epsilon}_t^k - \boldsymbol{\nabla}\mathbf{\Phi}^t \cdot \boldsymbol{\epsilon}_0^k + \delta\tilde{\mathbf{\Phi}}^t + \boldsymbol{\nabla}\mathbf{\Phi}^t \cdot (\mathbf{a}_0^k - \mathbf{x}_0)} \,, \tag{III.18}$$

using the same short notations as before. For the locally-constant, Eq. (III.8a) is modified to give:

$$\boxed{\boldsymbol{\mu}_{\mathrm{LC}} - \mathbf{x}_t \approx \sum_k \omega_k \boldsymbol{\epsilon}_t^k - \boldsymbol{\nabla}\mathbf{\Phi}^t \cdot \sum_k \omega_k \boldsymbol{\epsilon}_0^k + \delta\tilde{\mathbf{\Phi}}^t + \boldsymbol{\nabla}\mathbf{\Phi}^t \cdot (\boldsymbol{\mu}_0 - \mathbf{x}_0)} \,. \tag{III.19}$$

In this last equation, the noises $\boldsymbol{\epsilon}_0^k$ and $\boldsymbol{\epsilon}_t^k$ are averaged with weights ω_k. This averaging procedure gives new centered Gaussian variables, with a variance divided by $\approx K$ (to be more precise, the variance is divided by $\approx K/2$ if one uses the procedure of Lguensat et al. (2017)). This shows that there is a bias-variance trade-off: raising the value of K lowers the variance and raises the bias of the estimator of $\mathbf{x}_t$. For the locally-incremental, Eq. (III.9a) becomes:

$$\boxed{\boldsymbol{\mu}_{\mathrm{LI}} - \mathbf{x}_t \approx \sum_k \omega_k \boldsymbol{\epsilon}_t^k - \boldsymbol{\nabla}\mathbf{\Phi}^t \cdot \sum_k \omega_k \boldsymbol{\epsilon}_0^k + \delta\tilde{\mathbf{\Phi}}^t + \left(\boldsymbol{\nabla}\mathbf{\Phi}^t - \mathbf{I}\right) \cdot (\boldsymbol{\mu}_0 - \mathbf{x}_0)} \,. \tag{III.20}$$

Similar equations can be derived using the description with $\mathbf{f}$ instead of $\mathbf{\Phi}$. For the sake of brievety, these equations are not shown here.

For the locally-linear operator, it is not as straightforward to derive the modified form of Eq. (III.10), because the linear regression between the analogs and successors is modified due to the additive noises. The model equations (III.17a) and (III.17a) with $\delta = 0$ are known as error-in-variable models (Carroll et al., 2006). The locally-linear operator described in this chapter only assumes noise in the dependent variables, while

error-in-variable models assume noise in the independent variables, which is best suited to measurement error purposes. The correct hypothesis for applying the locally-linear operator would be recovered in the case $\boldsymbol{\epsilon}_0^k = \mathbf{0}$. Applying regular least-squares solution for a linear regression, when the real context is the one of measurement errors in the independent variables, is known to induce bias in the estimation of linear coefficients. In the one-dimensional, linear case, the estimation of linear coefficient is biased towards zero, but it is not so clear in the case of local-linear regression of a non-linear model (Griliches and Ringstad, 1970). Total least squares (Markovsky and Van Huffel, 2007) could be used to account for measurement errors in a more precise way then regular least-squares do. However, for the time being, we consider only direct application of the locally-linear operator, as presented before.

In the strong noise limit and if K is small, variations of $\mathbf{a}_t^k$ (respectively $\mathbf{a}_0^k$) with k are dominated by variations of $\boldsymbol{\epsilon}_t^k$ (respectively $\boldsymbol{\epsilon}_0^k$) and there is no correlation between successors and analogs, such that the least-squares estimate of the linear relationship $\mathbf{S}$ between the analogs and successors gives a matrix of zeros, and the locally-linear operator reduces to the locally-constant operator. However, if K is large enough the linear relationship between the analogs and successors can be found through averaging in the least-squares estimate.

Numerical experiments with the system of Lorenz (1963) were performed, identical to those produced in Fig. III.6, but with additive noises $\boldsymbol{\epsilon} \sim_{\text{i.i.d.}} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, and σ is a given percentage (0.01 or 1) of the root-mean-squared distance between two points picked randomly on the attractor, and with values of K equal either to 40 or 200. Results are presented in Fig. III.9. For strong noise and low K, the regression between analogs and successors is unable to capture the real flow's Jacobian. Raising the value of K makes the Jacobian estimation from analogs a little more accurate. In both cases (i.e. $K = 40$ or 200), the estimation is independent from the condition number of analogs, which is always low because the two-dimensional structure of the attractor is lost inside the isotropic, three-dimensional noise. For weak noise, the dependency of the estimation on the analogs' condition number is recovered, especially when raising the value of K. The poorest estimation of the Jacobian is thus witnessed for high condition numbers. However, even for low condition numbers, the estimation error is much larger than in the noise-free case of Fig. III.6.

Similar experiments are conducted on the same Lorenz system to produce forecasts whose mean errors are averaged over 1 000 points on the attractor. Results are presented
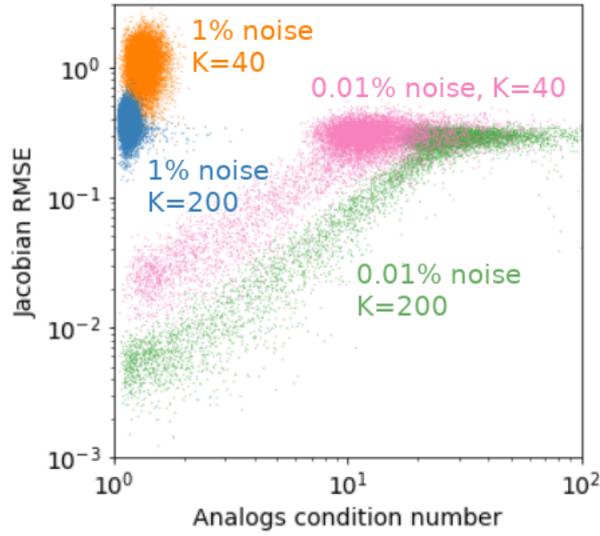
Figure III.9:   Same as Fig. III.6 but with additive noise. The experiment was repeated with low and high noises (0.01% and 1% of the RMS distance between two points on the attractor), and high and medium values of the number of analogs K (200 and 40).

in Fig. III.10.

First, let us consider the case of *high noise* (1% of RMS distance in the attractor). In this case, it can be seen that all methods (locally-constant, incremental and linear) give the same results. First, notice that in the case of large noise, $\mathbf{S} \approx 0$ and the locally-linear operator tends to the locally-constant operator. Then, for small lead times, the noise terms of Eq. (III.19) and Eq. (III.19) are dominant. This explains why raising the value of K diminishes the forecast error. Then, for large lead times, the terms proportional to $(\boldsymbol{\mu}_0 - \mathbf{x}_0)$ in Eq. (III.19) and Eq. (III.19) are dominant. Therefore, raising the value of K does not change much the accuracy of mean analog forecasting.

Then, consider the case of *low noise* (0.01% of RMS distance in the attractor). For small lead times, there is no dominant term in Eq. (III.19) and Eq. (III.19). Thus, raising the value of K diminishes the errors due to noise but also raises the errors proportional to $(\boldsymbol{\mu}_0 - \mathbf{x}_0)$, resulting in better performances for the locally-linear, but similar performances for the locally-constant and locally-incremental. For large lead times, all methods are equivalent (this property is independent of noise), and the terms proportional to $(\boldsymbol{\mu}_0 - \mathbf{x}_0)$ are dominant, such that raising the value of K increases bias and raises the mean analog forecasting errors.

To complete this analysis, the analytical expression of mean locally-linear error $\boldsymbol{\mu}_{\mathrm{LL}} - \mathbf{x}_t$

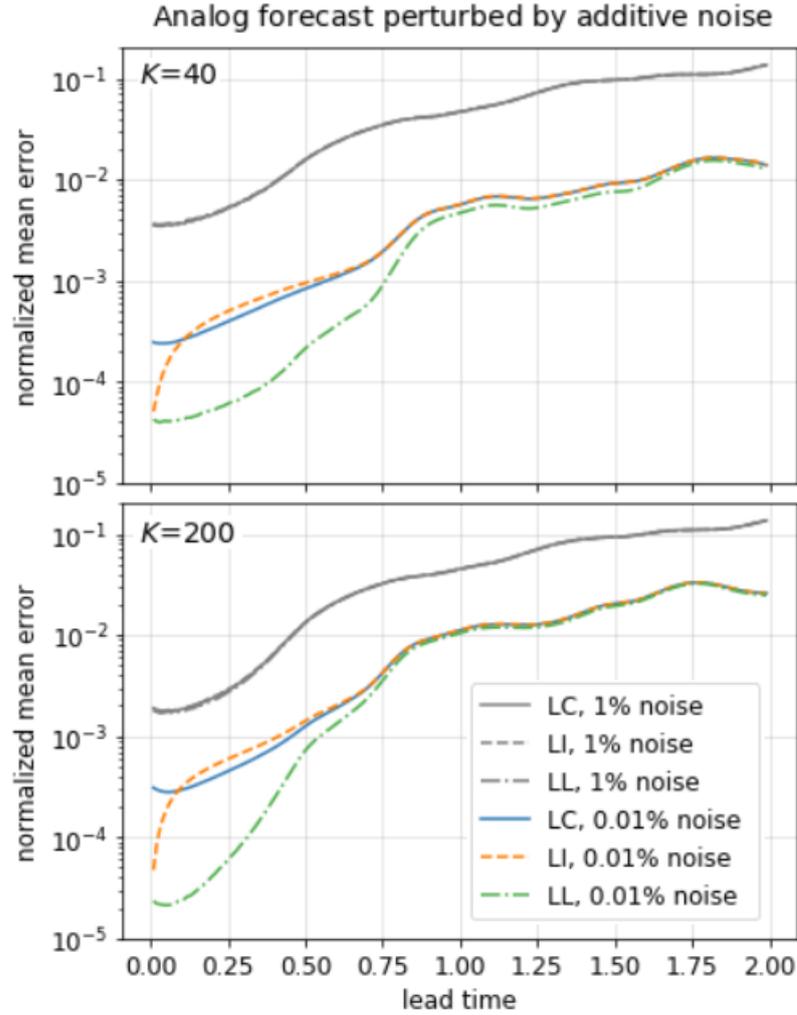Figure III.10: Influence of noise intensity σ and number of analogs K on analog forecasting operators, using the Lorenz (1963) system, a catalog size of $L = 10^7$ and a time step $\mathrm{d}t = 0.01$ between elements of the catalog. Average errors of mean analog forecasting operators $< \boldsymbol{\mu}_{\mathrm{L}.} - \mathbf{x}_t >$ are computed over 1 000 points taken randomly on the attractor.

is yet to be written in the case of additive noise.

3.3.2 Interpreting analog forecasting operators with an error-in-variables model

In this section, analytical expressions for the probability distribution of the future state $\mathbf{x}_t$ are given, assuming the error-in-variables model (III.17), exact linearity of the flow, and i.i.d. Gaussian observational noises. The Jacobian of the flow is seen as a model parameter, for which particular values allow to retrieve particular analog-forecasting operators.

Neglecting the difference between the analogs' flow and the real flow, and omitting the higher-order terms in Eq. (III.18), one finds:

$$\mathbf{a}_t^k - \mathbf{x}_t = \boldsymbol{\epsilon}_t^k - \boldsymbol{\nabla}\boldsymbol{\Phi}^t \cdot \boldsymbol{\epsilon}_0^k + \boldsymbol{\nabla}\boldsymbol{\Phi}^t \cdot (\mathbf{a}_0^k - \mathbf{x}_0)\,, \tag{III.21}$$

where the initial state $\mathbf{x}_0$ is known almost surely, the Jacobian $\boldsymbol{\nabla}\boldsymbol{\Phi}^t$ is assumed to be fixed (we will see later that it can be viewed as a parameter of the analysis). Therefore, the future state $\mathbf{x}_t$ is a random variable that is linked to the random noises $\boldsymbol{\epsilon}_0^k$, $\boldsymbol{\epsilon}_t^k$. Let us write $\hat{\mathbf{x}}^k := \mathbf{a}_t^k + \boldsymbol{\nabla}\boldsymbol{\Phi}^t \cdot (\mathbf{x}_0 - \mathbf{a}_0^k)$, then

$$\mathrm{d}\mathbb{P}(\mathbf{x}_t = \mathbf{x}) = \int \prod_{k=1}^{K} \left(\delta\left[\mathbf{x}_t - \hat{\mathbf{x}}^k + \boldsymbol{\epsilon}_t^k - \boldsymbol{\nabla}\boldsymbol{\Phi}^t \cdot \boldsymbol{\epsilon}_0^k \right] \mathrm{d}\mathbb{P}(\boldsymbol{\epsilon}_0^k) \, \mathrm{d}\mathbb{P}(\boldsymbol{\epsilon}_t^k) \right) \, \mathrm{d}\mathbf{x}\,,$$

where $\delta[\cdot]$ is the Dirac delta function. The integral is taken over the values of the $\boldsymbol{\epsilon}_0^k$ and $\boldsymbol{\epsilon}_t^k$ for all $k \leq K$, and it is assumed that the noises are independent. If one further assumes that the noises are identically distributed and follow a centered Normal distribution with covariance $\sigma^2 \mathbf{I}$ where $\sigma > 0$, $\mathbf{I}$ is the identity matrix of size n, and n is the dimension of the space in which $\mathbf{x}_t$ lives, then the integral can be evaluated analytically:

$$\mathrm{d}\mathbb{P}(\mathbf{x}_t = \mathbf{x}) = \left(\frac{1}{\sqrt{(2\pi\sigma^2)^n \det \mathbf{J}}} \right)^K \exp\left\{ -\frac{1}{2\sigma^2} \sum_{k=1}^{K} \left(\mathbf{x} - \hat{\mathbf{x}}^k \right)^{\mathrm{T}} \mathbf{J}^{-1} \left(\mathbf{x} - \hat{\mathbf{x}}^k \right) \right\} \mathrm{d}\mathbf{x}\,,$$

which is the product of Gaussian probability density functions of mean $\tilde{\mathbf{x}}^k$ and covariance matrix $\sigma^2 \mathbf{J}$ where $\mathbf{J} = \mathbf{I} + (\boldsymbol{\nabla}\boldsymbol{\Phi}^t)(\boldsymbol{\nabla}\boldsymbol{\Phi}^t)^{\mathrm{T}}$. It follows that the future state $\mathbf{x}_t$ is a Gaussian random variable with mean $\boldsymbol{\mu}_K$ and covariance $\boldsymbol{\Sigma}_K$:

$$\boxed{\boldsymbol{\mu}_K = \frac{1}{K}\sum_{k=1}^{K}\hat{\mathbf{x}}^k \;, \quad \boldsymbol{\Sigma}_K = \frac{\sigma^2}{K}\mathbf{J}}\;.$$

Setting $\boldsymbol{\nabla}\boldsymbol{\Phi}^t = 0$ implies that $\hat{\mathbf{x}}^k = \mathbf{a}_t^k$ and $\mathbf{J} = \mathbf{I}$. In this case, one recovers the expression for the mean of the locally-constant with uniform weights $\omega_k = \frac{1}{K}$. The covariance matrix is $\frac{\sigma^2}{K}\mathbf{I}$, such that the variability of the future state decreases when K grows (more information is added). The locally-constant operator presented in Chapter II estimates the covariance matrix of the random variable $\mathbf{x}_t$ as the empirical covariance of the successors, which should be $\approx \sigma^2\mathbf{I}$ if $\boldsymbol{\nabla}\boldsymbol{\Phi}^t = 0$, therefore, the locally-constant operator might overestimate the covariance of $\mathbf{x}_t$. However, this must be tempered by the fact that in practice, $\boldsymbol{\nabla}\boldsymbol{\Phi}^t$ is not zero, which adds variability to the process of making a forecast with a larger number of analogs K.

Setting $\boldsymbol{\nabla}\boldsymbol{\Phi}^t = \mathbf{I}$ implies that $\hat{\mathbf{x}}^k = \mathbf{x}_t + \mathbf{a}_t^k - \mathbf{a}_0^k$ and $\mathbf{J} = 2\mathbf{I}$. One then recovers the mean of the locally-incremental with uniform weights $\omega_k = \frac{1}{K}$. The covariance matrix is twice as large as in the case $\boldsymbol{\nabla}\boldsymbol{\Phi}^t = \mathbf{0}$. Same remarks can be given as to the difference between the theoretical covariance matrix and the empirical covariance matrix from the operator presented in this chapter.

This small calculation gives a new interpretation of the analog forecasting operators which was not given before in the literature. It motivates the use of Gaussian random variables to perform analog forecasting, as is done in Lguensat et al. (2017), combined with ensemble Kalman filtering techniques. This analysis is reminiscent of extended Kalman filtering techniques which deal with non-linear time-evolution models.

3.4 Analog short-term ocean wave forecasting from point-measurements

In this section, we show an application of analog forecasting to ocean wave time-series using wave buoy data. Analog forecasting is compared with an auto-regressive model that was reported as the most efficient method in another comparative study.

Wave energy converters (WEC) draw energy from wind-generated waves, i.e. the deformation of the ocean surface. To achieve optimal energy conversion, WEC are tuned depending on the wave conditions from spectral forecasts, but short-term deterministic forecasts ($\sim$20 s ahead) can add an appreciable gain. Note that other offshore renewable energy devices, such as offshore windmills, might also benefit from this forecast to optimize wind energy intake. To do so, one can rely on the past measurements of sea-surface eleva-

tion at the WEC location. Fusco and Ringwood (2010) review the forecasting strategies for this purpose, with tests on data from wave buoys. The paper does not mention analogs, and we show here how short-term wave forecast for energy converters might benefit from analog forecasting methods. Leads for further research will be given to complement the presented preliminary results.

Fusco and Ringwood (2010) report that auto-regressive (AR) models show the best performances for the forecast of ocean waves, with lead times of less than 20 seconds, and using only past wave records (tens of minutes) at the location where the future waves will be forecasted. Therefore, analog methods will be compared to AR models. To test the performances of analogs methods, point-wave measurements from a wave buoy installed in the Iroise sea will be used. The wave buoy is located south-west off the Sein island, at (48°00'20.5"N 4°55'20.1"W), where the water depth is ∼45 meters. Therefore, for small wave heights of $\sim 4m$ or less, the deep-water criterion is met, and the approximation of linear waves is justified. This is even more so true for forecast horizons below 20 seconds. Fusco and Ringwood (2010) show that AR models are justified physically by linear wave theory, such that the conditions for comparison between analog methods and AR methods with our dataset are advantageous.

The data is a single 30 minutes-sample of sea-surface elevation measurement from the buoy, with acquisition frequency f_a =1.25Hz (every Δt =0.8 s). The wave spectrum from this sample is shown in the left panel of Fig. III.11. It corresponds to what is described by Fusco and Ringwood as a wide-banded sample, for which AR performances are not 100% satisfying for lead times approaching 20 seconds. Indeed, for narrow-banded sea-states, AR forecasts capture the future state with almost 100% accuracy for all lead times below 20 seconds. To mimic the test conditions of Fusco and Ringwood, the sea-surface elevation signal was low-pass filtered (see the right panel of Fig. III.11). Then, the sample is divided into training and testing sets. The first 20 minutes will be used for analogs and successors (training), and the last 10 minutes will be used to perform forecast and compare with the truth (testing). The filtered sea-surface elevation at time t is noted η_t.

Time-embeddings (Takens, 1981) are used to define the state at time t, with a time lag of 1 acquisition and n delayed-coordinates. Thus, analogs of the vector $(\eta_{t-(n-1)\Delta t}, \ldots, \eta_{t-\Delta t}, \eta_t)$ are sought for, and successors η_{t_k+l} are used to forecast η_{t+l}, where t_k is the time of the k-th analog's last element. Results are shown in Fig. III.12. The locally-incremental gave results similar to the locally-constant, and is therefore not shown. The locally-constant (lower-left panel of Fig. III.12) is applied directly: the state at time $t + l$ is
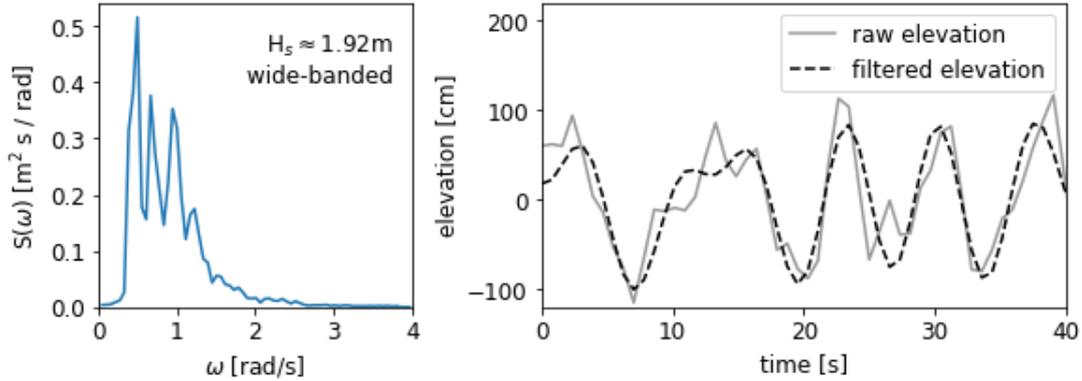
Figure III.11: Left: wave spectrum from 30 minutes-sample buoy sea-surface elevation data. Right: raw and low-pass filtered elevation, with a rough cutoff at frequency $\sim$1.1 rad/s.

forecasted from the state at time t (t is observation time and l is forecast horizon) using the locally-constant operator once (the forecast is not split into multiple forecasts). The locally-linear, on the other hand, is applied both directly (lower-right panel of Fig. III.12) and using multiple forecasts as explained in the Sec. **??** (middle-right panel of Fig. III.12). In this multiple forecast version, the state at time $t+l$ is predicted from the state at time t using the locally-linear operator $l \times f_a$ times in a row. The AR method performs a linear regression between the vectors $(\eta_{t-(p+n-1)\Delta t}, \ldots, \eta_{t-(p+1)\Delta t}, \eta_{t-p\Delta t})$ and $\eta_{t-(p-1)\Delta t}$, for $p \in [1, N]$, where N is the size of the training window for the AR method. The number of analogs used is set to $K = 200$, and the size of the training window for the AR model is also $N = 200$. Here, AR and multiple LL are very similar, only the training dataset changes.

Results of these experiments are reported in Fig. III.12. From the example forecast (top figure), one sees that although the performances of the locally-constant operator are not as good as other techniques (but still not bad compared to other techniques from Fusco and Ringwood 2010), this operator seems to be capturing some low-frequency tendencies. The AR, direct LL and multiple LL have perfect performances for small lead times (or forecast horizon). The multiple LL outperforms the AR for large lead times, and these operators are both outperformed by the direct LL with $n = 12$. Similar plots with larger values of n did not yield better results. In particular, the AR and multiple LL exhibit instability for large n and large lead times (this can be witnessed at the end of the $n = 10$ curve for the AR).

Looking now at statistical performances (bottom plots of Fig. III.12), the multiple LL and AR methods are very similar, which explains that the curves for $n = 6$ and $n = 8$ are almost identical. The AR performs only one optimization of regression coefficients on the N last observations, and uses these coefficients recursively for all forecasts. This poses stability issues: if the optimized AR model has eigenvalues with modulus > 1, the forecast will diverge after some time. Contrarily, the multiple LL performs a different optimization for each forecast over the K analogs, such that instabilities will be damped by the varying values of the regression coefficients. Also, the LL uses information of the LC forecast as basis for the linear regression, which reduces slightly the forecast error. Finally, the direct LL is even more stable as it performs only one linear forecast, such that n can be raised to a higher value (here $n =$12), allowing for even better performances.

The fact that the LL forecasts are working using analogs of the past 30 minutes is justified physically by the stability of sea-state conditions. An AR model is stable over a few hours according to Fusco and Ringwood (2010), such that there is no need to compute the regression coefficients over the last observations, and one can use observations further in the past.

To go beyond this first investigation, these tests should be tried again on other buoy data samples to check the consistency of the results. Also, information on the horizontal position of the buoy (north-south and east-west) could be used to define analogs, possibly raising accuracy. Then, one could try to use analogs from past sea-states (other days but same buoy, for instance) and using a rescaling strategy. Perhaps, this would allow even for the locally-constant to reach high accuracy, by raising the number of good analogs. Finally, an interesting lead would be to test the performances of the schemes in the case of extreme waves. Having a robust forecast scheme for extreme events may allow for the WEC to go in safety mode and avoid getting damages from large waves.

An example forecast of a rogue wave from the method introduced earlier is shown in Fig. III.13. The direct locally-linear forecast with $n = 14$ gives the best average performances on this sample (based on a figure similar to Fig. III.12, not shown), and the forecast of the rogue wave with this analog method is compared with the true sea-surface elevation. The forecasts 10 seconds ahead and 5 seconds ahead of the rogue wave are shown. With 10 seconds of lead time, the crest height of the wave is greatly underestimated, although the predicted wave is still moderately high (of probability inferior to 0.05). The rogue wave is not included in the 97.5% confidence interval of the analog forecast. With 5 seconds of lead time, the rogue wave is well estimated, but the following
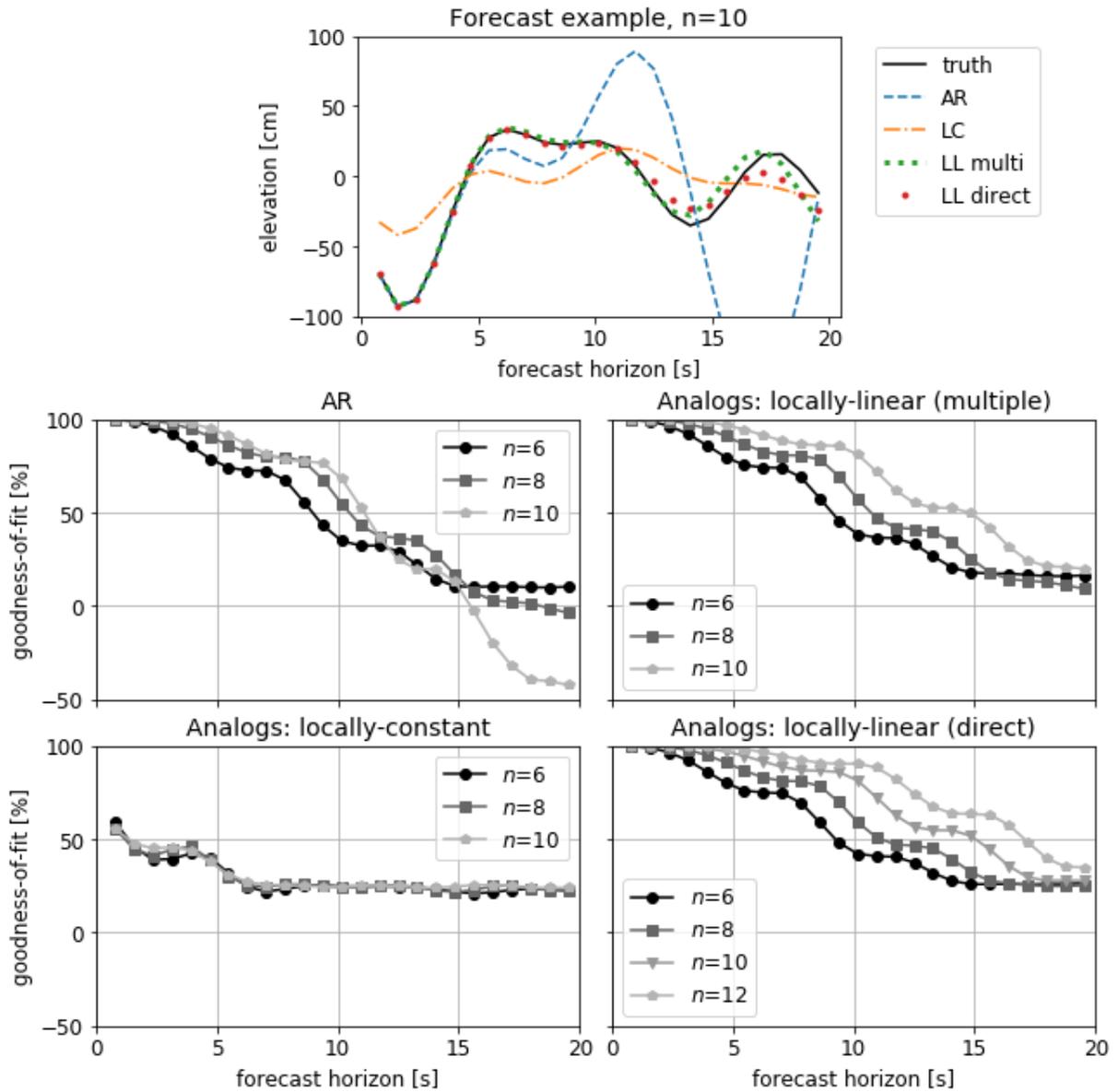
Figure III.12: Results of different forecasting strategy on ocean wave point-measurements from 30-minutes buoy sample described in Fig. III.11. The number of time-delays (or length of time-window) is n.
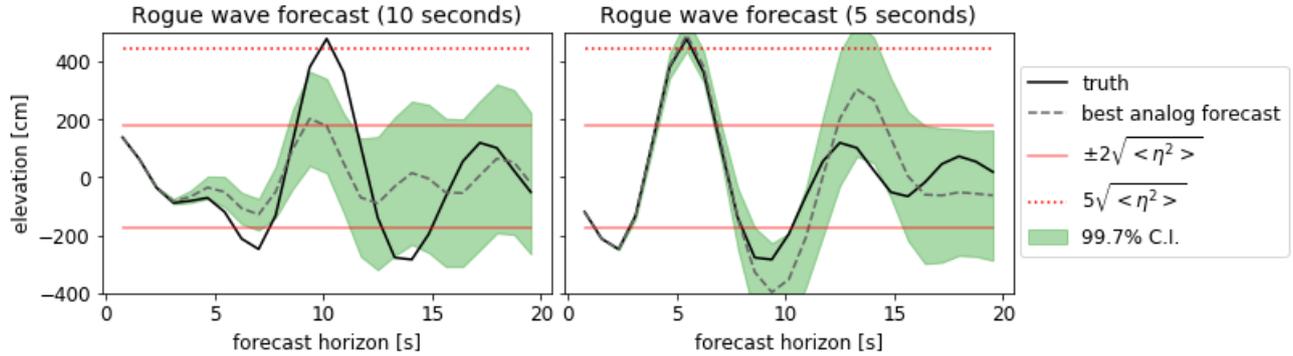
Figure III.13: Forecasting a rogue wave from the best (in terms of averaged error) analog forecasting strategy, which here is the locally-linear with $n = 14$ time-delays. The average forecast (gray, dashed) and 99.7% confidence intervals (green) assuming Gaussian margins are shown. The rogue wave threshold for sea-surface elevation is five times the standard deviation (red, dotted). A lower threshold of $\pm$ two standard deviations is shown (red, full), which corresponds to an average probability of less than 0.05 assuming Gaussian margins. The standard deviation is calculated from 20 minutes of observation preceding the rogue wave.

wave heights are overestimated, although included in the (large) confidence interval. Note that these accuracies are consistent with average statistics for the forecasting method: at 5s lead time, the goodness-of-fit is of $\approx 90\%$, and drops to $\approx 45\%$ for lead times of 10s (not shown). The fact that the proposed methodology is still working for this rogue wave indicates that non-linear effects are negligible, although wave buoys are known to bias observations of ocean waves towards linear waves. This observation is coherent with the fact that this large wave has a profile reminiscent of Sleppian profiles: for linear waves, the mean profile around a maxima is given by the auto-correlation function (Phillips et al., 1993). Note that auto-regressive forecasts are based on auto-correlation too. If one is concerned with the occurrence of very large waves, then estimating the probability of very large events is crucial, and criterion based on average forecast error might not be optimal. Here it is seen that the locally-linear operator, although efficient in terms of average error, seems to struggle in estimating the probability of occurrence if this large wave. To assess this more precisely, forecasts over a large number of rogue waves would have to be performed. From a physical point of view, the theory of Phillips et al. (1993), which gives the mean and variance of profiles around a maximum crest height, could be used to produce statistical forecasts of rogue waves.

This last example motivates the design of forecasting strategies tailored to extreme

waves. Taking a different point of view, the next chapter focuses on the forecast of large amplitude ocean waves from one-dimensional space-time wave fields, while the last examples used only time-samples. The possibility to measure the waves away from the point of interest greatly increases the forecast efficiency for lead times way beyond the ones considered here.

4 Summary

☐ The locally-constant and locally-linear operator (Lguensat et al., 2017) have a mean forecast error that is linear in average analog-to-initial state distance. The locally-incremental has the advantage that its forecast error tends to zero for small lead times.

☐ The locally-linear operator has a mean forecast error that is quadratic in analog-to-initial state distance. It is able to estimate a projection on the local attractor of the flow's Jacobian matrix.

☐ In terms of mean accuracy and for small lead times, the locally-linear is superior to the locally-incremental, which is superior to the locally-constant. For large lead times, the three operators have similar performances. Note that the locally-linear has a higher numerical cost and requires a larger number of analogs because it performs a linear regression.

☐ The locally-constant and locally-incremental can be seen as particular cases of the locally-linear operator where the flow's Jacobian matrix is replaced by the zero matrix (i.e., the flow is constant) for the locally-constant and by the identity matrix (i.e., the flow is a constant plus the identity map) for the locally-incremental.

☐ If there is a difference between the analogs' flow and the real system's flow (i.e., a bias in observational dynamics), this difference increases the forecast error. At first order (i.e., if the difference is small enough), this increase is *additive*.

☐ Analog forecasting performances are decreased by the presence of additive, centered observational noise in the catalog of analogs and successors. However, the aforementioned results remain valid for noises of small amplitude. Our theoretical arguments and numerical experiments indicate that the use of a larger number of

analogs decreases the influence of centered additive noise, in particular for small lead times.

☐ It can be interesting to split a forecast to a large lead time into multiple analog forecasts of smaller individual lead times. This is particularly true when using the locally-linear operator, for which our numerical experiments show that this splitting procedure allows to gain of orders of magnitude in precision. The maximal lead time at which the splitted forecasts remain consistent is twice as high as in the case of direct forecast. However, our wave forecast experiments show that this splitting procedure can also decrease stability. The applicability of this method appears to be case-dependent.

☐ The process of splitting into multiple forecasts of smaller lead times motivates the comparison between analog forecasting and finite-difference numerical schemes. The latter approach the solution of differential equations in the limit of small time-step (or equivalently, of large number of time-grid points), whereas analog forecasting approaches the solution in the limit of large catalog size. Using this comparison and results from the previous chapter, we link the convergence rate of analog forecasting to the attractor dimension.

Further research would have to focus on the case of partial observations. A more detailed investigation of successor spread would be interesting because the latter is important for ensemble analog forecasts.

FORECAST OF EXTREME OCEAN WAVES FROM CREST VELOCITIES

Until the rainbow burns the stars out in the sky (always)
Until the ocean covers every mountain high (always)...

– Stevie Wonder, *As*

1 Introducing the article

In the previous chapter, the empirical short-term forecast of incoming ocean waves with analogs for lead times up to 20 seconds was considered. This type of forecast can allow to optimize energy intake from offshore renewable energy devices. However, one might also be interested in forecasts of extreme events such as the so-called "rogue" waves. Indeed, due to safety issues, the installation, maintenance and normal functioning of offshore (including renewable energy) devices are all impossible when subject to extreme waves, winds or rain. The present chapter focuses on the forecast of extreme waves from crest velocities, which could be extracted from the measurement of wave fields through X-band radars, possibly combined with stereo video cameras. The short-term forecast of large waves could have two major consequences for offshore installations. First, the applicability range of offshore operations could be raised. If the largest amplitude waves can be predicted to turn on a "safety mode" before their arrival, then more severe sea states can be considered for maintenance or installation. Second, the safety of structures can be increased by the ability to turn on safety mode in the case of incoming large amplitude waves, even when offshore operations are not considered. We investigate the following question:

- Can high-amplitude "rogue" ocean waves be predicted from wave crest (and trough) velocities?

To answer this question, the following strategies and tools are used:

∗ Large-amplitude ocean waves are assumed to arise from the linear focusing of narrow-banded, unidirectional surface gravity waves. In this context, we identify the largest future waves with the maximal focusing stage of wave packets.

∗ We derive analytical expressions for the crest velocities of Gaussian wave packets. This allows to establish a simple relationship between spatial variations of crest velocities and the focusing or defocusing nature of a given wave packet. In particular, the time evolution (or "profile") of an approximate crest velocity "gradient" is derived. Based on a fitting of crest velocity measurements to this profile, the future position, horizontal width, and relative amplitude (with respect to present) of a focusing Gaussian wave packet are estimated.

∗ We use numerical experiments of linear ocean waves to assess the validity of the results and the applicability of the method to linear Gaussian wave packets.

∗ The validity of our method for second-order non-linear wave packets is discussed using physical arguments.

∗ The robustness of our method to higher-order non-linear effects is gauged from numerical experiments of the non-linear Schrödinger equation.

2 Article published in *Natural Hazards*: "Wave group focusing in the ocean: estimations using crest velocities and a Gaussian linear model"

Paul Platzer[1][2][3][4] • Jean-François Filipot[4] • Philippe Naveau[1,2]
Pierre Tandeo[3] • Pascal Yiou[1,2]   doi:10.1007/s11069-020-04279-z

1. LSCE, CEA Saclay l'Orme des Merisiers, UMR 8212 CEA-CNRS-UVSQ, U Paris-Saclay, IPSL, Gif-sur-Yvette, France
2. ESTIMR - Extrèmes : Statistiques,Impacts et Régionalisation, LSCE, Gif-sur-Yvette, France
3. IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238, Plouzané, France
4. France Énergies Marines, Plouzané, France

2.1 Abstract

Wave group focusing gives rise to the formation of large gravity waves at the surface of the ocean, some of which are called rogue waves and represent a natural hazard for ships and offshore platforms. For safety purposes, it is crucial to predict when and where these large waves will appear and how large they will be. This work focuses on crest velocities, a quantity that is relatively easy to extract from sea-surface elevation fields. It is shown that there is a direct link between crest velocity gradient and wave group linear dispersive focusing. Studying analytically the focusing of one-dimensional Gaussian wave packets under linear evolution makes it possible to derive estimates of quantities at focus, based only on crest velocity measurements. In this way the focusing time, focusing size and focusing amplitude (relative to instantaneous amplitude) of an isolated Gaussian wave packet can be estimated. Our work is also applicable to second-order non-linear waves. Limitations due to higher-order non-linear effects are studied in numerical simulations of the non-linear Schrödinger equation.

2.2 Introduction

Human and material losses caused by large amplitude ocean surface waves are significant, as reported by Kharif and Pelinovsky (2003). The term "rogue wave" refers to waves with a crest-to-trough wave height larger than $2H_s$, where H_s is the significant wave height and is defined as four times the standard deviation of the sea surface elevation Dysthe et al. (2008). The threshold of $2H_s$ is arbitrary, and in the following "rogue wave" might simply refer to waves with a large height compared to local statistics. Many rogue waves have been measured in the ocean, including the famous 26m-high Draupner wave (Haver, 2004). Even in moderate sea states, the occurrence of rogue waves can be problematic for operations (e.g. cable layout from a ship) or for transferring staff from a ship to an offshore platform or wind turbine. The harmful potential of rogue waves is a natural motivation for the forecast of their position, amplitude and spatial width over time, based on real-time measurements available from instruments that can be mounted on ships or offshore platforms. The short-term forecast of high-amplitude ocean waves must neither be confused with tsunami forecast (Titov et al., 2005), nor sea-state forecast (Niclasen et al., 2010), nor short-term forecast of sea-surface elevation for wave energy converters (Fusco and Ringwood, 2010).

Many physical processes could be involved in the formation and growth of oceanic

rogue waves (Adcock and Taylor, 2014). In this paper, we focus on the formation of large surface gravity waves in deep water due to dispersive focusing, a process that we briefly describe here. In deep water, the sea surface elevation can be approximated by an infinite sum of plane waves propagating at a velocity given by the dispersion relation of ocean surface gravity waves (Longuet-Higgins, 1957). This approximation is based on the linearization of the physical equations at the ocean surface, assuming small steepness of the sea surface elevation. Dispersion causes waves of different wavelengths to propagate at different velocities, which creates mixing. When this mixing results in the constructive interference of a large number of waves (superposition of their crests, or troughs), this leads to a wave of large amplitude, and we call this process dispersive focusing. Here, we use a linear description of long-crested (one-dimensional in space) ocean waves, and we tackle the issue of forecasting the dispersive focusing of ocean wave packets based on crest velocities. Thus, the proposed methodology is not limited to waves exceeding the rogue threshold of $2H_s$, but aims at forecasting the maximum amplitude of any wave packet under dispersive focusing.

Other processes can lead to the focusing of ocean waves, such as wave-current interactions (Peregrine, 1976; Touboul and Kharif, 2016; Quilfen et al., 2018). Also, the focusing of ocean waves can increase wave steepness, leading to second-order non-linear effects that act on the shape and height of waves (Forristall, 2000), or even higher-order non-linearities potentially leading to the instability of Benjamin and Feir (1967). However, we ignore wave-current interactions, and first investigate linear dispersive focusing. The latter appears to be a necessary first step in the formation of the largest ocean surface gravity waves (Fedele, 2008) and is thought to be the main mechanism driving the formation and evolution of real-life rogue waves, along with second-order non-linearities (Fedele et al., 2016). Although designed for linear waves, our work is applicable to second-order non-linear waves, and we also examine the limitations of our study due to higher-order non-linear effects.

A review of the possible strategies for rogue wave forecasting in the absence of currents and based on only space-time measurements of the sea-surface elevation can be found in Slunyaev (2017). For a review and comparison of numerical methods for deterministic short-term prediction of water wave fields, the reader is referred to Klein et al. (2020), which also lists commercial decision support systems based on X-band radar acquisition. However, the forecast of high-amplitude waves does not require prediction of the whole wave field, and forecast methods may be targeted only at high-amplitude waves. Slunyaev

et al. (2005) developed the idea that rogue wave formation could be inferred from energy flux convergence. A predictive scheme based on such a principle was developed by Guo and Alam (2017). However, the method laid out in Guo and Alam (2017) requires knowledge about the field of fluid velocity in the whole water column, a quantity that is difficult to measure in real time. Slunyaev (2017) proposed that the energy transfer could be inferred from the local group velocity, an idea which is closely related to what we propose here.

The envelope of the wave field is not well defined when the wave spectrum is not narrow enough. For this reason, our method relies on the velocities of crests and troughs, which are theoretically accessible from fields of sea-surface elevation. Today however, in-situ ocean wave measurement technology needs to improve in order to achieve the forecasting of large ocean waves a few minutes in advance, independently of the method that is to be used. The combination of X-band radar images (Borge et al., 2013) with stereo video cameras (Benetazzo et al., 2018; Filipot et al., 2019) is promising, and motivated us to focus on crest velocities. For the same applications, the sea surface reconstruction using lidar illumination is another encouraging alternative (Nouguier et al., 2013).

Fig. IV.1 shows the basic principle of the proposed methodology. A wave packet under linear evolution in a dispersive medium (where the phase velocity is frequency-dependent) is either in the stage of focusing (a, d), is focused (b, e), or is in the stage of defocusing (c, f), and those three stages always take place in this order. When the wave packet is focused, the velocities of the crests and troughs (maxima and minima) are nearly homogeneous (e), and smaller than the phase velocity corresponding to the peak frequency of the Gaussian Wave Packet (GWP). This slow-down effect is described in the linear wave theory outlined by Fedele (2014). Conversely, when the wave packet is not focused, crest and trough velocities are neither uniform in space, nor constant in time (d, f): there is a strong gradient of crest velocity (or, to be more precise, a gradient of specular velocity, see Sect. 2.3, subsection 2.3.3). The sign of this gradient reveals the focusing or defocusing nature of the wave packet, while its absolute value is proportional to the growth rate of the wave packet amplitude (for narrow-banded linear wave packets). By tracking the evolution of this gradient over time, it is possible to predict the time at which the GWP is focused, the size of the GWP at focus, and the ratio between the amplitude at focus and the instantaneous amplitude.

Theoretical background and hypotheses are outlined in Sect. 2.3, with details of derivations given in the appendix. The method is tested in numerical experiments of linear waves in Sect. 2.4. In Sect. 2.5 we investigate the applicability of the method to

weakly non-linear wave packets using numerical experiments of the non-linear Schrödinger equation.

2.3 Theory for linear Gaussian wave packets

2.3.1 Definitions and hypotheses

We refer to Gaussian Wave Packets (GWP) as one-dimensional sea-surface elevation profiles η_G such that the modulus of the space Fourier transform of η_G is a Gaussian function of wavenumber k. Such packets were widely used to generate water waves in laboratory experiments (Clauss and Bergmann, 1986), and to study theoretically and numerically the dispersive formation of large amplitude waves (for instance, see Pelinovsky et al., 2011). The Gaussian shape assumption allows us to carry detailed calculations for the crest velocities. We can write η_G as the real part of its complex analytic representation Ψ_G with

$$\Psi_G(x,t) = \frac{A_{\mathrm f} L_{\mathrm f}}{\sqrt{2\pi}} \int \exp\left\{ -\frac{L_{\mathrm f}^2 (k-k_0)^2}{2} + i\left(kx - \Omega(k)t + \Phi(k)\right) \right\} \mathrm{d}k, \qquad \text{(IV.1)}$$

where x is position, t is time, $A_{\mathrm f}$ is the amplitude of the GWP at focus, $L_{\mathrm f}$ is the spatial extent of the GWP at focus (also further referred to as "group size"), k_0 is the peak wavenumber of the GWP, $\Omega(k)$ is the angular frequency given by the dispersion relation, and $\Phi(k)$ is the phase of the GWP. All quantities are real unless otherwise stated.

We make the narrow-band approximation that $L_{\mathrm f} k_0$ is large, which induces relative errors that are all proportional to a negative power of $L_{\mathrm f} k_0$. The effects of surface current, wind input and wave dissipation are ignored. Non-linearities are first neglected, but their effect is studied in the last section.

2.3.2 Approximation of Ψ_G

Let us write

$$\Omega(k) = \omega_0 + c_g(k-k_0) + \frac{c^*}{2k_0}(k-k_0)^2 + \text{h.o.t.}, \quad \Phi(k) = \phi_0 - x_0(k-k_0) + \frac{S}{2}(k-k_0)^2 + \text{h.o.t.},$$

where $x_0 = -\frac{\mathrm{d}\Phi}{\mathrm{d}k}(k_0)$ and $S = \frac{\mathrm{d}^2\Phi}{\mathrm{d}k^2}(k_0)$. We can then write Ψ_G as
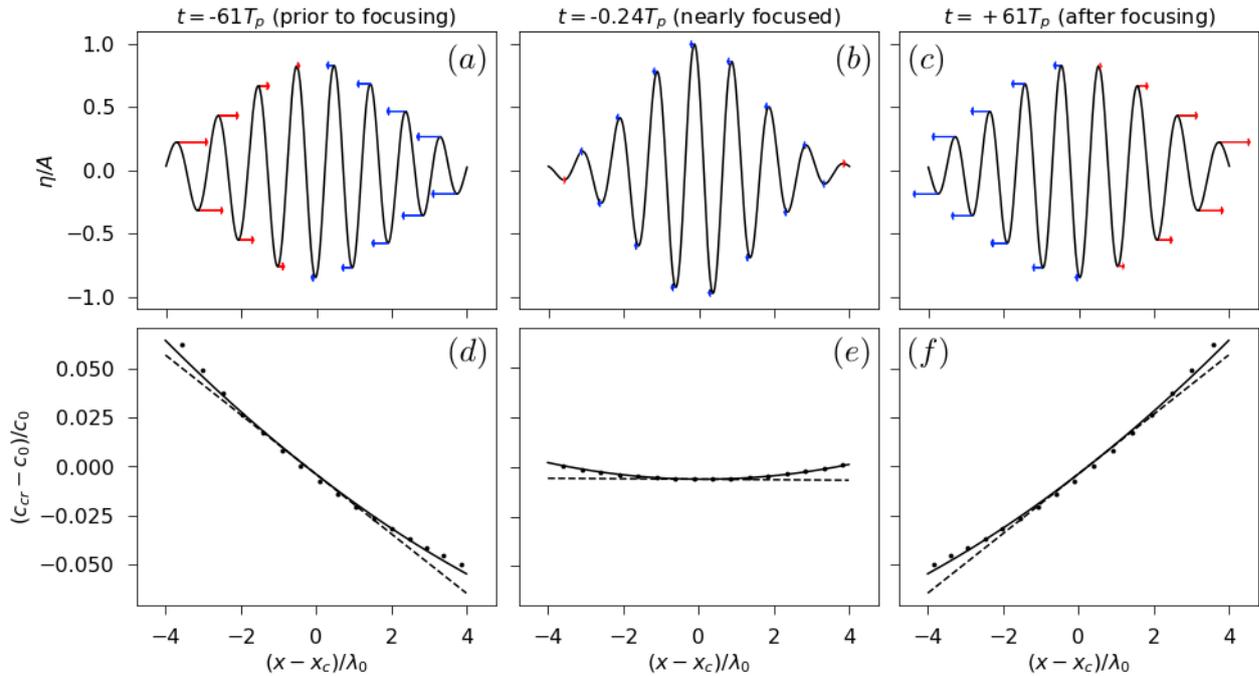
Figure IV.1: Crest and trough velocities of a Gaussian wave packet (GWP) at different times under linear evolution. The GWP is moving from left to right, and we assume the dispersion relation of deep-water gravity waves. The peak phase velocity is noted c_0. Extracted from the numerical experiments described in Sect. 2.4, with $L_f k_0 = 10$. Abscissa : position relative to the center of the GWP normalized by the peak wavelength $\lambda_0 = 2\pi/k_0$. Black curves in $(a-c)$ are a numerical approximation of η_G calculated using Eq. (IV.1). Arrows in $(a-c)$ have lengths proportional to $c_{cr} - c_0$, red when $c_{cr} > c_0$ and blue when $c_{cr} < c_0$. Black dashed lines in $(d-f)$ are a first-order approximation of the analytical prediction c_{sp}^a. Black curves in $(d-f)$ are second-order approximation. Black dots in $(d-f)$ are crest velocities c_{cr} extracted from numerical experiment. Shown are the situations prior to focusing (a and d), nearly focused (b and e), after focusing (c and f).

$$\Psi_G = \frac{A_f L_f}{\sqrt{2\pi}} e^{i(k_0 x - \omega_0 t + \phi_0)} \int \exp\left(f(k - k_0) + \text{h.o.t.}\right) \mathrm{d}(k - k_0) ,$$

where

$$\forall K, \ f(K) = -0.5 \left[L_f^2 + i\left(\frac{tc^*}{k_0} - S\right) \right] K^2 + i\left[x - c_g t - x_0 \right] K .$$

Since $L_f k_0 \gg 1$ we can neglect higher-order terms in the exponential and integrate with respect to $k - k_0$. This is a specific case of Laplace's method. Rearranging, this gives

$$\Psi_G^a(x,t) = A(t) \exp\left\{ \frac{-(x - x_c(t))^2}{2L(t)^2} + i\left(k_0 x - \omega_0 t + \phi(x,t)\right) \right\}, \qquad \text{(IV.2)}$$

where the a-superscript stands for "approximation" and with:

$$A^4(t) = \frac{A_f^4}{1 + (c^* t / k_0 L_f^2)^2} ,$$

$$L^2(t) = L_f^2 + \left(\frac{\frac{c^* t}{k_0} - S}{L_f} \right)^2 ,$$

$$\phi(x,t) = -\arctan\left(\frac{\frac{c^* t}{k_0} - S}{2L_f^2} \right) + \frac{(x - x_c(t))^2}{2L^2 L_f^2} \left(\frac{c^* t}{k_0} - S \right) ,$$

and $x_c(t) = x_0 + c_g t$. Setting $t \to t - S k_0 / c^*$, and $x \to x - x_0 + S k_0 c_g / c^*$, we obtain the simpler expressions :

$$\frac{A(t)}{A_f} = \left(1 + (t/\tau)^2\right)^{-1/4} , \qquad \frac{L(t)}{L_f} = \left(1 + (t/\tau)^2\right)^{1/2} , \qquad \text{(IV.3)}$$

$$\phi(x,t) = \phi_0 - \arctan(t/2\tau) + (t/2\tau) \frac{(x - x_c)^2}{L(t)^2}, \qquad \text{(IV.4)}$$

where $x_c = c_g t$ is the center of the GWP, and $\tau = k_0 L_f^2 / c^*$ is the group contraction time scale (possibly negative). From our hypotheses we have $|\tau| \ll \omega_0^{-1}$ so that $A(t)$, $L(t)$ and $\phi(t)$ are slowly-varying compared to the carrier plane wave $e^{i(k_0 x - \omega_0 t)}$. See Fig. IV.2 for an illustration of these quantities.

Using the work of Olver (1968), one can show that Ψ_G^a gives an approximation of Ψ_G / A_f that is valid up to the second order in $1/(k_0 L_f)$, the next order being the fourth one. It is also clear from the work of Olver (1968) that obtaining higher-order approxi-
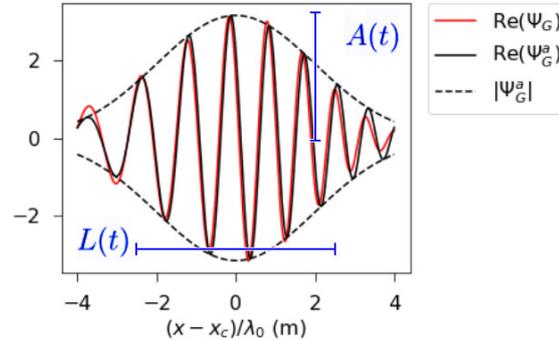
Figure IV.2: A Gaussian wave packet a few periods before focusing, with related notations. The approximation Φ_G^a and Φ_G are almost identical, the largest discrepancies are found at the tails of the GWP. This GWP has parameter $L_f k_0 = 5$ and $A_f = 5$.

mations of the right-hand side of Eq. (IV.1) would require information on the 3rd and higher-order derivatives of $\Phi(k)$ at k_0, that could not be set to zero by changing the origin of the time and space axis.

The expression of Ψ_G^a was given by Kinsman (1965) for the case $\Phi(k) = 0$. In Eq. (IV.2), we generalize to the case where $\Phi(k)$ is a slowly varying function of k.

2.3.3 Crest velocity approximation

In one dimension, we can follow the trajectory of a crest $x_{cr}(t)$ by imposing

$$\partial_{xx}\eta\left(x_{cr}(t_0), 0\right) < 0 \ , \forall t > t_0, \quad \partial_x \eta\left(x_{cr}(t), t\right) = 0 \ ,$$

This implies that $c_{cr}(t) = c_{sp}(x_{cr}(t), t)$ where $c_{sp} = -\frac{\partial_{xt}\eta}{\partial_{xx}\eta}$ is the velocity of points of constant slope called specular points (Longuet-Higgins, 1957). Specular points would be indicated to a distant observer as the points where light was reflected from a distant source. The velocity of specular points can be defined at every point of the surface and contains the crest velocities. For identical reasons, trough velocities can also be calculated using the velocity of specular points. Using Eq. (IV.2) we can express c_{sp}^a (an approximate form of c_{sp} based on Ψ_G^a) as a function of $X \equiv k_0(x - x_c)$ and $T \equiv t/\tau$:

$$c_{sp}^a(X, T) = \frac{C_1(T)X^4 + C_2(T)X^3 + C_3(T)X^2 + C_4(T)X + C_5(T)}{C_6(T)X^3 + C_7(T)X^2 + C_8(T)X + C_9(T)}$$

with the $C_i(T)$ being time-dependent coefficients which all have terms proportional to the small parameter $(L(t)k_0)^{-2}$. For the sake of brevity, we do not give the expressions

of those coefficients, but only the first terms of the Taylor expansion of $c_{\mathrm{sp}}^{\mathrm{a}}$ with respect to $(L(t)k_0)^{-2}$. We also assume the dispersion relation of deep-water gravity waves (i.e. $\Omega \propto \sqrt{k}$)

$$\frac{c_{\mathrm{sp}}^{\mathrm{a}}(X,T)}{c_0} = 1 + \frac{5}{8}\frac{[-1-0.8XT]}{(L(t)k_0)^2} + \frac{5}{8}\frac{[1+XT+(0.2+T^2)X^2]}{(L(t)k_0)^4} + \mathcal{O}\left(\frac{1}{(L(t)k_0)^6}\right), \;\; \text{(IV.5)}$$

where $c_0 = \omega_0/k_0$. Neglected higher-order terms are of $\mathcal{O}\left((Lk_0)^{-6}\right)$, and can only become large when $|x - x_c| \gg L(t)$ or $|t/\tau| \gg 1$, cases where the height of the waves is negligible. As pointed out by one of the reviewers and as mentioned in the previous section, the fourth-order terms in this expression have been derived from Ψ_G^a which is a second-order approximation of Ψ_G. For this reason, they cannot be trusted in theory. However, the numerical experiments shown in this paper suggests that they give a reasonable approximation of c_{sp}.

At zeroeth order $c_{\mathrm{cr}}^{\mathrm{a}}$ equals c_0, which corresponds to the monochromatic case. At first order it is a time-dependent linear function of X (see the dashed lines of Fig. IV.1($d-f$)). At second order it is a time-dependent second-order polynomial in X (see the full curves of Fig. IV.1($d-f$)). Good agreement is found between $c_{\mathrm{cr}}^{\mathrm{a}}$ which was derived from the approximate Ψ_G^{a}, and c_{cr} which is associated with Ψ_G (see the dots of Fig. IV.1($d-f$)). Other experiments with lower values of $k_0 L_f$ also show a good agreement between $c_{\mathrm{cr}}^{\mathrm{a}}$ from Eq. (IV.5) and c_{cr} (not shown).

Note that Eq. (IV.5) is valid only for specular points of slope zero $\partial_x \eta = 0$, which is the case for crests and troughs. The general expression for c_{sp}^a is more complicated than Eq. (IV.5), but by abuse of notation and for our purpose c_{sp}^a denotes the right-hand side of Eq. (IV.5), which is a Eulerian quantity while c_{cr} is Lagrangian. In particular, this allows us to compute derivatives of the right-hand side of Eq. (IV.5).

2.4 Numerical experiments of linear waves

2.4.1 Numerical set-up

We performed simulations based on a numerical approximation of the integral in Eq. (IV.1). All parameters were fixed except $L_{\mathrm{f}}k_0$. We assumed the dispersion relation of deep-water gravity waves $\Omega(k) = \sqrt{gk}$ where g is the acceleration of gravity at the earth's surface, set to 9.81 ms^{-2}. Φ was set to be a constant without loss of generality up to a

change of reference frame, so that the GWP is focused at $t = 0$. Setting $A_f = 5$ m and $k_0 = 0.02$ rad m^{-1} gives $c_0 \approx 22.15$ ms^{-1} and $T_p = 2\pi/\sqrt{gk_0} \approx 14$s. Our choice for k_0 corresponds to typical dominant surface gravity waves encountered in the ocean. We did numerical experiments with three different values for $L_f k_0$, either $L_f k_0 = 2.36$ or $L_f k_0 = 5$ or $L_f k_0 = 10$.

In all experiments, the integral was approximated using 150 wavenumbers, ranging from $k_{\min} = 0.001$ rad m^{-1} to $k_{\max} = 0.05$ rad m^{-1} when $L_f k_0 = 2.36$; from $k_{\min} = 0.005$ rad m^{-1} to $k_{\max} = 0.037$ rad m^{-1} when $L_f k_0 = 5$; and from 0.012 rad m^{-1} to 0.028 rad m^{-1} when $L_f k_0 = 10$. Further refinement in the wavenumber discretization did not yield significant changes in the results. The integral defining $\Psi_G(x, t)$ was evaluated over a space-time grid with constant steps dx $= 3.14$ mm and dt $= 1.17$ s when $L_f k_0 = 2.36$; dx $= 2.55$ mm and dt $= 1.36$ s when $L_f k_0 = 5$; dx $= 3.37$ mm and dt $= 1.56$ s when $L_f k_0 = 10$. In all experiments, $(x - x_c)$ varies from $-2\lambda_0$ to $+2\lambda_0$. t varies from $-4|\tau| \approx 14T_p$ to 0 when $L_f k_0 = 2.36$; from $-2|\tau| \approx 32T_p$ to 0 when $L_f k_0 = 5$; and from $-|\tau| \approx 64T_p$ to 0 when $L_f k_0 = 10$. The time step was chosen to be comparable with X-band radar typical measurement frequency, which are on the order of 1Hz. Note that high spatial resolutions allow the precise measurement of small crest velocity variations (see the vertical axis of Fig. IV.1($d - f$)) for convenience in testing the method. In practice, time-averaging techniques could be used to lower spatial resolution constraints. Each experiment was run a thousand times for different values of the constant phase Φ, picked randomly between -π and π.

2.4.2 The choice of $L_f k_0$

The typical spectral shape of individual wave packets in the ocean is unknown, and thus our justification for the choices of values for $L_f k_0$ will only be partial. Isolated wave packets only exist conceptually, and in the ocean, one wave packet is always at least partially merged with other neighboring packets. Ocean wave measurements allow for the reliable estimation of the spectrum of a sample of the sea-surface. Such a sample is made of many packets merged together. We will call such a spectrum a "total spectrum" to make a clear distinction with the spectrum of one individual packet. To our knowledge, there is no theory which gives the conditional probability that a measured wave packet has a given spectrum (corresponding to A_f, L_f and k_0 in our case) assuming that the total spectrum is known. Longuet-Higgins (1984) links the temporal width of wave packets (linked to our $L(t)k_0$) with the width of the total spectrum ν. But the probability of

measuring a packet with an instantaneous size $L(t)k_0$ is different from the probability for a given packet to have a focusing size $L_f k_0$. However, a wave packet that is far from focusing will have a large spatial width but a small amplitude, and it will thus have a lower probability of being measured at sea as it would be merged with other packets of greater amplitude.

Cousins and Sapsis (2015b) performed numerical simulations of ocean waves with Gaussian spectra and randomly drawn phases, coupled with an algorithm for the detection of the spatial width and amplitude of individual wave packets. From those simulations, the authors were able to compute probability distributions for a given wave packet to be observed at any point in time, with a given spatial width and amplitude. In those numerical simulations, the packets observed with an amplitude larger then the rogue wave threshold $A(t) > H_s$ had a spatial width $L(t)k_0$ close to the inverse of the width of the total spectrum. As those packets had large amplitudes, it is likely that they were focused or close to focusing. This seems to indicate that for a given spectrum, the packets which have a high amplitude are likely to have a spectral width comparable to that of the total spectrum.

The case $L_f k_0 = 2.36$ corresponds to a spectral width of 0.15, $L_f k_0 = 5$ corresponds to 0.071, and $L_f k_0 = 10$ corresponds to 0.035. Recall that we use the definition of spectral width through moments of the temporal frequency : $\nu^2 = m_0 m_2 / m_1^2 - 1$ and $m_i \propto \int S(f) f^i \mathrm{d}f$ for all i with f the temporal frequency and $S(f)$ the frequency spectrum, i.e. the density of energy per frequency. In the ocean, the most dangerous waves are due to large values of H_s, associated with storm seas. In these sea states, the largest individual waves are the so-called dominant waves that contain the energy of a narrow spectral band around the peak frequency, f_p. Banner et al. (2000) suggested that this spectral band ranges from $0.7 f_p$ to $1.3 f_p$. Computing ν over this spectral band for storm wave spectra computed at a deep water point in the North East Atlantic ([-4.5°W, 46°N]) during the entire 2013-14 winter and using the WaveWatchIII spectral wave model (Tolman et al., 2009) leads to values lying between 0.05 and 0.15. We also note that the directional spreading for this particular spectral band is generally limited (about 20° using the same dataset). Our numerical experiments thus show how our method should behave for dominant waves in typical storm conditions.
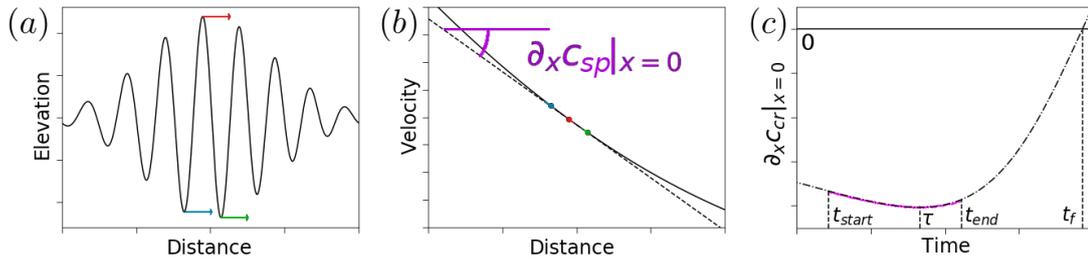
Figure IV.3: Step-by-step methodology. (a) Crest detection and velocity estimation. (b) Velocity gradient estimation. (c) Velocity gradient estimation over time, fitted and extrapolated.

2.4.3 Validity of the narrow-band approximation for $c_{\mathrm{sp}}^{\mathrm{a}}$

We know from our analytical derivations based on Ψ_G^{a} that the specular velocity gradient at the center of the GWP contains information on the focusing properties of the packet, and we also know that this gradient is almost homogeneous in the spatial dimension x. Thus, we hope to estimate this gradient based on measurements of crest and trough velocities around the center of the GWP. We start by measuring the crest velocities of the three crests/troughs that are closest to the center of the GWP and the distances between them, giving us three values of $c_{\mathrm{sp}}(\cdot, t)$ (see Fig. IV.3(a)). We chose to use only three values because we want this procedure to be applicable to any wave packet inside a random unidimensional field (meaning one space dimension and one time dimension) of linear surface gravity waves containing many different packets. In this case the crests and troughs of a given wave packet that are far away from its center may be mixed up with crests and troughs from other packets. In the case of a random unidimensional field of linear surface gravity waves, our procedure would be equivalent to finding the largest crest or trough in the packet and measuring its velocity and those of is two neighboring troughs or crests, respectively.

Then, at any time t we make a polynomial fit of second order in space from our three values of $c_{\mathrm{sp}}(\cdot, t)$ (full curve of Fig. IV.3(b)), and we keep the linear part of this fit as our approximation for $\partial_x c_{\mathrm{sp}}|_{x=x_c}$ (dashed curve of Fig. IV.3(b)).

We can compare the result of this procedure (Fig. IV.3(a, b)) to our approximate analytical prediction $\partial_x c_{\mathrm{sp}}^{\mathrm{a}}|_{x=x_c}$ that we obtain from Eq. (IV.5) when neglecting terms of $\mathcal{O}\left((Lk_0)^{-6}\right)$. This is plotted in Fig. IV.4. In the case where $L_{\mathrm{f}} k_0 = 2.36$, it can be seen that our procedure causes oscillations in the estimation of $\partial_x c_{\mathrm{sp}}|_{x=x_c}$. This is due to the fact that the phase velocity of surface gravity waves is twice as large as their group
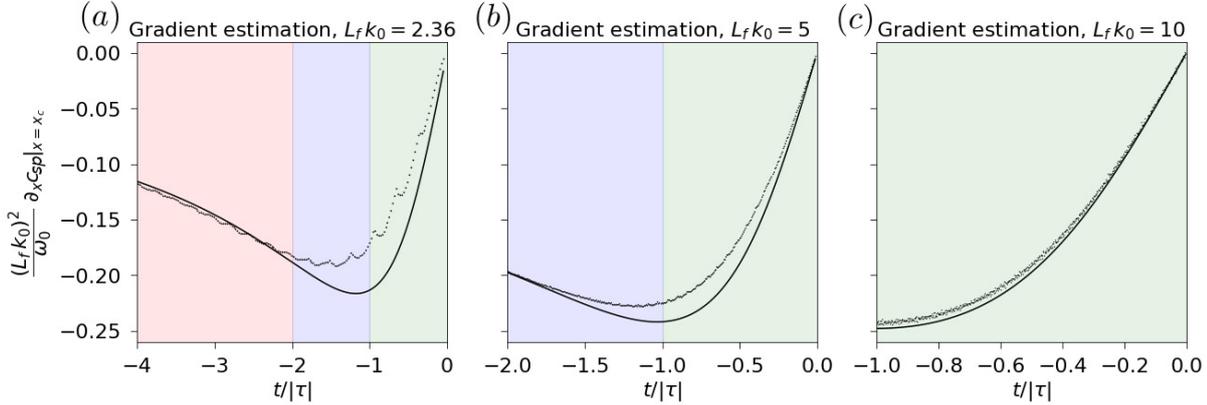
Figure IV.4: Normalized crest velocity gradient at the center of the GWP. The full curves correspond to the second order analytical prediction of $\partial_x c_{cr}^a|_{x=x_c}$ from Eq. (IV.6). The dots are approximations of $\partial_x c_{cr}$ from three crest/trough velocities around $x = x_c$ from the numerical experiments. Green background is for $-|\tau| \leq t \leq 0$, violet for $-2|\tau| \leq t \leq -|\tau|$, and pink for $-4|\tau| \leq t \leq -2|\tau|$.

velocity, and thus crests move twice as fast as the center of the wave packet. Thus every $\Delta t \approx 2T_p$, the crest or trough in the front of the GWP that we use for our approximation of $\partial_x c_{sp}|_{x=x_c}$ is no longer one of the three closest crests or troughs from the center of the GWP, and it is thus replaced by another trough or crest, respectively, coming from the back of the GWP. These oscillations still exist for larger values of $L_f k_0$, but they appear faster at the time scale $|\tau|$, since we have $|\tau|/T_p = 2(L_f k_0)^2/\pi$ from the definitions in Sect. 2.3.2. As $L_f k_0$ decreases, our approximations are less valid, especially around $t = -|\tau|$, but the behaviour of our measured $\partial_x c_{sp}|_{x=x_c}$ is still coherent with $\partial_x c_{sp}^a|_{x=x_c}$ and the approximations are relevant when t is far enough from $-|\tau|$. It can be observed that for $L_f k_0 = 5$ and 10, the minimum of our approximation for $\partial_x c_{sp}^a|_{x=x_c}$ is achieved close to $t = -|\tau|$. This is because the term in parentheses in Eq. (IV.6) is close to 1. On the other hand, for $L_f k_0 = 2.36$, we can see that the minimum of $\partial_x c_{sp}^a|_{x=x_c}$ is achieved slightly before $t = -|\tau|$, and is of lower relative amplitude.

2.4.4   Estimation of focusing time, amplitude and width

Once we have an approximate measure of $\partial_x c_{sp}|_{x=x_c}$ over time between t_{start} and t_{end}, we want to find relevant information related to the focusing of the wave packet, such as t_f, $A_f/A(t_{end})$, and L_f. To do this, we fit our approximate measure of $\partial_x c_{sp}|_{x=x_c}$ over time to a known profile derived from Eq. (IV.5), which has the following form

$$\partial_x c_{\mathrm{sp}}^{\mathrm{a}}|_{x=x_c}(t) \approx \frac{2(t-t_{\mathrm{f}})}{\tau^2 + (t-t_{\mathrm{f}})^2} \left(1 - \frac{13/\omega_0|\tau|}{1 + ((t-t_{\mathrm{f}})/\tau)^2} \right). \tag{IV.6}$$

The fitting step corresponds to Fig. IV.3(c), and the right-hand side of Eq. (IV.6) is plotted in Fig. IV.4. We assume that ω_0 is known, thus the only parameters to be estimated are $|\tau|$ and t_{f}. These parameters can be read on the graph in Fig. IV.3(c) as t_{f} is at the crossing with the $y = 0$ horizontal line and τ is approximately at the minimum of the curve. Then, using $|\tau|$, t_{f} and Eq. (IV.3), we can estimate $A_{\mathrm{f}}/A(t_{\mathrm{end}})$, and further using ω_0 we estimate $L_{\mathrm{f}} = \sqrt{|\tau|\omega_0}/k_0$. In Fig. IV.5, we show the results of this procedure for the estimation of t_{f}/T_p, $A_{\mathrm{f}}/A(t_{\mathrm{end}})$ and $L_{\mathrm{f}}k_0$, for our three values of $L_{\mathrm{f}}k_0$ and for $t_{\mathrm{start}} = -4|\tau|; -2|\tau|; -|\tau|$ and t_{end} going from t_{start} to $t_{\mathrm{f}} = 0$. When $t_{\mathrm{end}} - t_{\mathrm{start}}$ is small compared to $|\tau|$, large variations in the values of the fitting parameters are observed. This variability is the consequence of variations in the estimation of $\partial_x c_{\mathrm{sp}}|_{x=x_c}$ which happen at the small time scale T_p, such as the oscillations described in Fig. IV.4. When $t_{\mathrm{end}} - t_{\mathrm{start}}$ is only a few T_p, those small-time-scale variations are incorrectly fitted, as if they corresponded to the profile from Eq. (IV.6), which varies at the large time-scale $|\tau|$. This causes incorrect estimations of the parameters t_{f} and τ. Systematic bias can be observed when $L_{\mathrm{f}}k_0 = 5$ and 10, with an over-estimation of focusing amplitude and group size, and underestimation of focusing time. This is consistent with the systematic bias observed in Fig. IV.4. It might be argued that for warning purposes, an error of less than 10% might not be critical, and thus that the systematic bias observed for $L_{\mathrm{f}}k_0 = 5$ and 10 should not be cause for concern. Estimations in the case where $L_{\mathrm{f}}k_0 = 2.36$ are challenging, because deviations from the hypothesis on which our method relies start to be large. However, our estimated values are still acceptable, which indicates a certain robustness in our method. An interesting fact is that as $L_{\mathrm{f}}k_0$ decreases, relative errors increase, but when coming back to the typical scales of ocean waves, these errors are of comparable amplitude. See for instance the green curve in Fig. IV.5(h), with a systematic bias of $\approx 2\%$, which gives an error of $\approx 0.1k_0^{-1}$ for L_{f}; similarly the green curve in Fig. IV.5(i), with a bias of $\approx 0.5\%$, gives an error of $\approx 0.05k_0^{-1}$ for L_{f}. Estimations of t_{f} give biases of the order of T_p, while those biases would seem much lower for $L_{\mathrm{f}}k_0 = 10$ if we had plotted estimates of t_{f}/τ rather than t_{f}/T_p.

As the time of contraction scales with τ (and thus with L_{f}^2), focusing properties of large groups (or groups of narrow spectra) can be predicted accurately a long time in advance, but they also require that the group crest velocity gradient be tracked far in advance. We
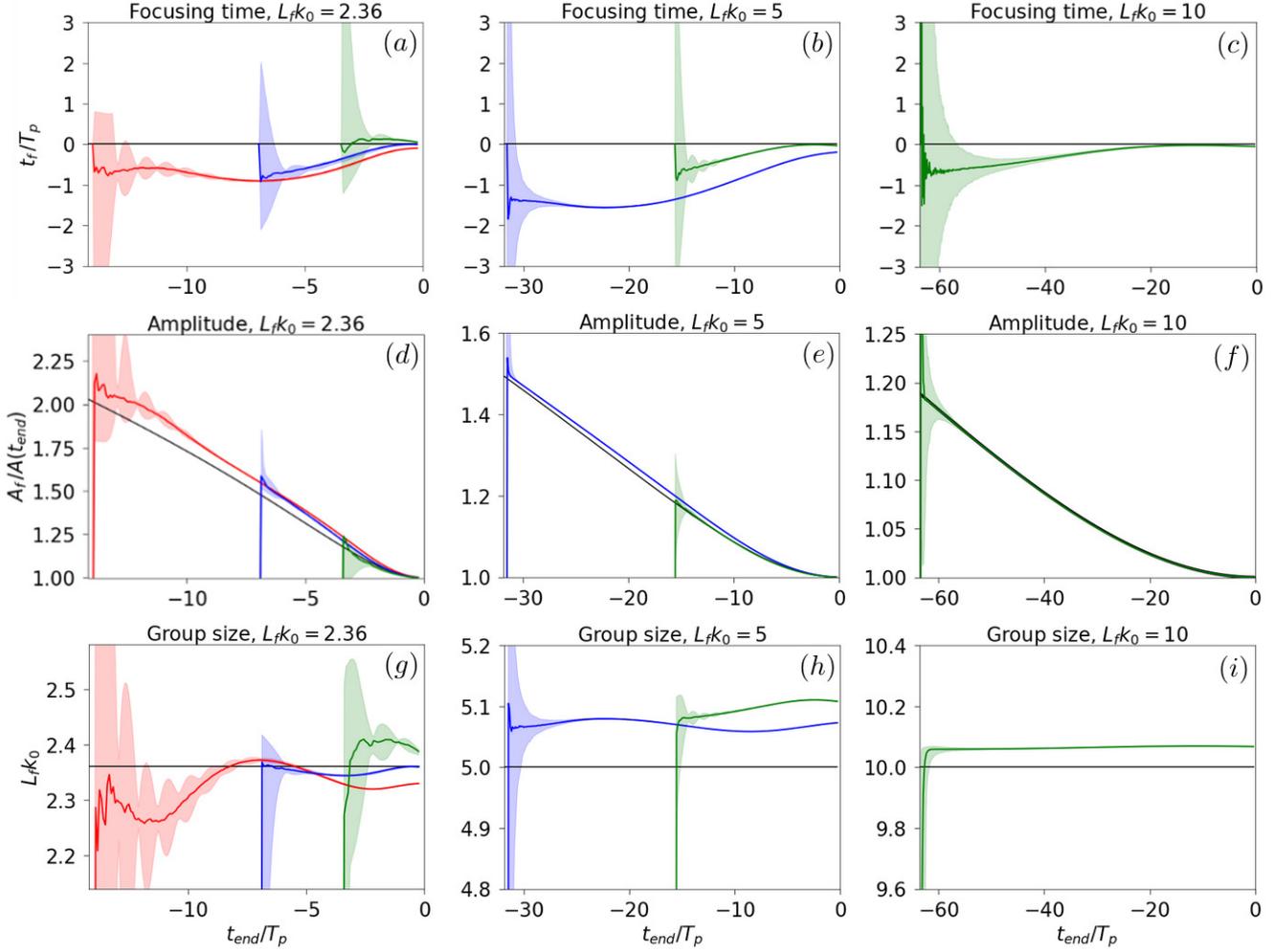
Figure IV.5: Estimation of focusing time t_f $(a - c)$, ratio of focusing amplitude and instantaneous amplitude $A_\mathrm{f}/A(t_\mathrm{end})$ $(d - f)$, and group focusing size L_f $(g - i)$, following the methodology described in Sect. 2.4.4. The objective (truth) is represented in each subplot by a black line. Estimations are based on approximate measurements of $\partial_x c_\mathrm{sp}|_{x=x_c}$ between t_start and t_end, with $t_\mathrm{start} = -4|\tau|$ (red) or $t_\mathrm{start} = -2|\tau|$ (blue) or $t_\mathrm{start} = -|\tau|$ (green). Experiments are repeated with different values of the phase Φ drawn randomly between $-\pi$ and π. The medians of predictions are shown in full lines and points between the 2.5 and 97.5 percentiles are shown in light shading. $L_\mathrm{f}k_0 = 2.36$ for (a), (d) and (g); $L_\mathrm{f}k_0 = 5$ for (b), (e) and (h); and $L_\mathrm{f}k_0 = 10$ for (c), (f) and (i). Vertical scales of $(g - i)$ are proportional to the value of Lfk_0.

note that it is technically challenging to predict the focusing of a wave packet 60 T_p in advance, because it requires having information on the wave field in a radius of about 10 km (for a typical peak period of 10 s) around the point of interest.

In our experiments we assumed that ω_0 is known. If ω_0 was unknown, we still could have used the lower order approximation $\partial_x c_{\text{sp}}^{\text{a}}|_{x=x_c}(t) \approx 2(t - t_{\text{f}})/(\tau^2 + (t - t_{\text{f}})^2)$, which does not depend on ω_0, and use it to evaluate t_{f} and $A_{\text{f}}/A(t_{\text{end}})$. The impact of such an approximation should not be strong in cases where $L_{\text{f}}k_0 = 5$ and 10, as we showed that including higher-order terms did not yield a significant change in the profile (Fig. IV.4). Then, however, no estimation of L_{f} could have been done.

2.5 Non-linear effects

2.5.1 General statements

Our work is based on the assumption of linear waves corresponding to infinitely small amplitudes $k_0 A_{\text{f}} \to 0$. Yet, real ocean waves have a finite amplitude and are thus non-linear. We examine here the consequences of non-linear effects on the validity of our method.

Second-order non-linearity

Second-order non-linearity affects the shape and amplitude of ocean waves. It makes crests sharper and troughs shallower, and raises the heights of crests and troughs by a quantity that is proportional to the product of the linear wave amplitude and the wave steepness (Forristall, 2000). Second-order non-linear effects modify the wave spectrum, although the latter is still time-independent. These effects modify neither the crest-to-trough wave height, nor the position of crests and troughs.

Here we use crest and trough velocities to predict GWP focusing time, focusing amplitude and focusing width. All the quantities mentioned in the last sentence are identical for linear and second-order non-linear waves. Therefore, our method is as valid for second-order non-linear waves as for linear waves. This is important because, for linear waves, the dispersion relation of surface gravity waves allows for very simple forecasting of the whole sea-surface using Fourier transforms. But this forecasting strategy cannot be applied to second-order non-linear waves for which the dispersion relation does not hold. The prevalence of second-order non-linear effects in the ocean thus further highlights the relevance of our method for real ocean waves. Note however that, from a practical point of view, the change in wave shape due to second-order non-linearity might make the accurate

measurement of trough positions more challenging.

Higher-order non-linearity

Higher-order (third-order and above) non-linear effects induce changes in shape, amplitude, and width of ocean wave packets. They make wave packet contraction stronger in the principal direction of propagation and weaker in the orthogonal direction (Adcock et al., 2012). Higher-order non-linearities can create surface gravity waves of very high amplitudes compared to linear waves, and have thus been studied extensively in the literature on rogue ocean waves. The spectrum of higher-order non-linear ocean waves is time-dependent, so the focusing time of higher-order non-linear wave packets is different from that of linear ones.

All these elements indicate that our forecasting strategy based on the linear wave hypothesis will fail to accurately forecast the focusing time, amplitude and width of a higher-order non-linear wave packet. However, we expect the quantities we want to predict to be continuous functions of the non-dimensional amplitude $k_0 A_f$ which controls the degree of non-linearity of the wave packet. Therefore, our method should still provide interesting results for finite but moderate values of $k_0 A_f$.

Following this preliminary analysis, we investigate the effects of higher-order non-linearities through numerical experiments of the one-dimensional non-linear Schrödinger equation (NLSE).

2.5.2   Numerical resolution of the NLSE

The one-dimensional NLSE, first described by Zakharov (1968), is a simple model for the evolution of the envelope of narrow-banded non-linear wave groups in deep water. Although it can model non-linearities above the second order, it is only valid for waves of moderate amplitude (called weakly non-linear waves). This is not a problem for us, as our concern is investigating the effects of higher-order linearities of moderate strength on our method. Comparison between solutions of the NLSE and fully non-linear simulations of water waves can be found in Henderson et al. (1999). In the reference frame moving at the group velocity, the NLSE can be written as follows:

$$i\partial_t u = \frac{\omega_0}{8k_0^2}\partial_x^2 u + \frac{1}{2}\omega_0 k_0^2 |u|^2 u \,, \tag{IV.7}$$

where $u(x,t)$ is the envelope of the wave field at position x and time t, ω_0 is the peak pulsation and k_0 the peak wavenumber. In the reference frame moving at the group velocity, the

envelope u and the wave field η are linked by the relation $\eta(x,t) = \text{Re}\left\{u(x,t)\exp i\left(\frac{1}{2}\omega_0 t - k_0 x\right)\right\}$ where $\text{Re}\{z\}$ denotes the real part of the complex number z.

We use a numerical set-up very similar to the one of Adcock and Taylor (2009). We solve the NLSE using a fourth-order Runge-Kutta scheme in time and a pseudo-spectral method in space. We set a peak period of 12 s and use a time step of 0.14 s. We use 2^{12} Fourier modes and a spatial discretization of 10m. All the simulations described here involve the following first steps:

- Take a linear GWP at focus with amplitude A_f^L and width $L_f^L = 5/k_0$ (we now add the ^{L}-superscript to discriminate linear and non-linear quantities).

- Back-propagate the GWP according to linear evolution for 65 time periods.

- Use this back-propagated GWP as an initialization and solve the NLSE numerically until the GWP is focused. Identify the focusing time t_f when the spatial maximum of wave amplitude $\max_x(|u(x,t)|)$ reaches a maximum. Measure A_f^{NL} and L_f^{NL}, the amplitude and width of the wave packet at focus, by fitting a Gaussian to the modulus of the envelope $|u(x,t_f)|$.

- Perform a cubic spline interpolation of u in space down to 3.3 cm resolution to allow for a precise measurement of crest velocities.

2.5.3 Non-linear crest velocity profiles

From those simulations, we measure $\partial_x c_{sp}|_{x=0}$ as in Sect. 2.4.3, with one modification to the method. This time, we measure the gradient only when one crest or one trough is in the middle of the wave packet such that $|x| \leq 4\lambda_0 dt/(2T_p)$, giving 4 measurements of the gradient around the position $x = 0$ every time period. By averaging those 4 measurements of $\partial_x c_{sp}$, we get a better estimate of $\partial_x c_{sp}|_{x=0}$, and we avoid the oscillations observed on the dotted lines in Fig. IV.4. Finally, to draw a complete profile of $\partial_x c_{sp}|_{x=0}$ over time, we repeat this procedure for different values of the constant and uniform phase.

Fig. IV.6 shows specular velocity gradient (approximated from crest and trough velocity measurements) versus time for different degrees of non-linearity. Focusing is always set at $t = 0$. Non-linear effects amplify the focusing, leading to a higher ratio A_f^{NL}/A_f^L and thus a lower ratio L_f^{NL}/L_f^L. This strengthening of the spatial concentration of energy is consistent with the higher values of $\partial_x c_{sp}|_{x=0}$ measured in Fig. IV.6 for higher values $k_0 A_f^{NL}$. The minimum of $\partial_x c_{sp}|_{x=0}$ is reached at $t = -|\tau|$ in the linear case, and it is reached for values of t closer to 0 for non-linear wave packets.
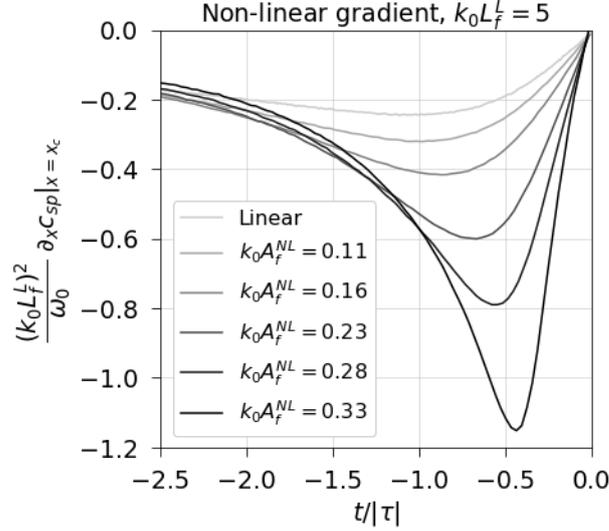
Figure IV.6: Phase velocity gradient estimated from crest velocities of non-linear GWP, as a function of time. The non-linear evolution of one-dimensional GWP is approximated through numerical simulations of the NLSE.

Numerical experiments for $k_0 L_{\mathrm{f}}^{\mathrm{L}} = 10$ (not shown here) indicate that the shape of the non-dimensional profile $\frac{(k_0 L_{\mathrm{f}}^{\mathrm{L}})^2}{\omega_0} \partial_x c_{\mathrm{sp}}|_{x=0}(t/|\tau|)$ depends both on $k_0 A_{\mathrm{f}}^{\mathrm{L}}$ and $k_0 L_{\mathrm{f}}^{\mathrm{L}}$ when higher-order non-linear effects are accounted for. As we have seen in Sect. 2.3, this profile was independent of $k_0 A_{\mathrm{f}}^{\mathrm{L}}$ and $k_0 L_{\mathrm{f}}^{\mathrm{L}}$ in the linear, narrow-banded case (low enough $k_0 A_{\mathrm{f}}^{\mathrm{L}}$ and high enough $k_0 L_{\mathrm{f}}^{\mathrm{L}}$).

We then evaluate the spatial profile of non-linear specular velocity at the time when the measured gradient $\partial_x c_{\mathrm{sp}}|_{x=0}$ reaches a minimum value. This is plotted in Fig. IV.7. We see that the quasi-quadratic spatial profile of specular velocity from Eq.(IV.6) is not valid for non-linear wave packets. For weakly non-linear GWP ($k_0 A_{\mathrm{f}}^{\mathrm{NL}} \leq 0.16$), we see that the gradient at $x = 0$ is higher than in the linear case. For even higher values of $k_0 A_{\mathrm{f}}^{\mathrm{NL}}$, the spatial scale of variations of c_{sp} is too small to estimate its spatial profile based on measurements of crest velocities which are separated by one wavelength in space.

A theoretical study of phase velocities for weakly non-linear wave fields is given in section VII of Yuen and Lake (1982). Figures 43, 44 and 45 show the departure from the linear waves relation $c_{\mathrm{p}}(k) \propto k^{-1/2}$. For various spectra, the study shows that the phase velocity ratio $c_{\mathrm{p}}(k)/c_{\mathrm{p}}(k_0)$ is augmented for wavenumbers both above and below the peak. In our focusing GWP case, shorter waves are in the front (downstream) and longer waves in the back (upstream), and both phase velocities will be augmented due to non-
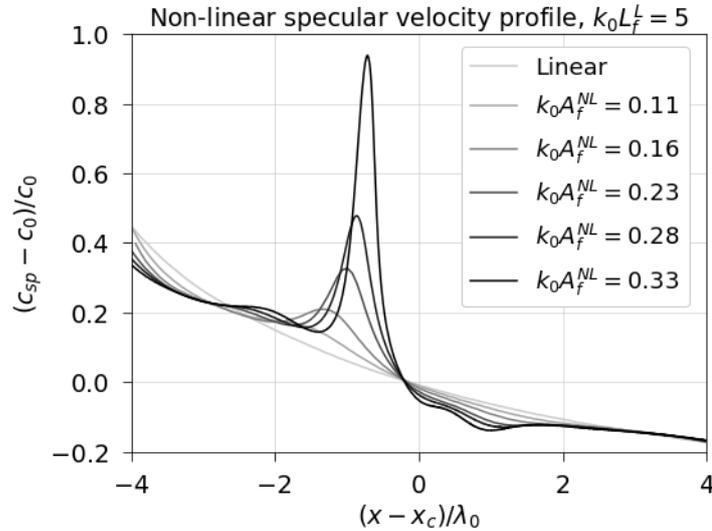
Figure IV.7: Crest velocity of non-linear one-dimensional GWP, as a function of the space coordinate x. The profile is estimated at the time when the central gradient of phase velocity is the lowest (see Fig. IV.6). The non-linear evolution of one-dimensional GWP is approximated through numerical simulations of the NLSE.

linear terms. To conclude this analysis, one would have to use information on the spatial distribution of wavenumbers in a weakly non-linear focusing wave packet. However, the study of Yuen and Lake (1982) is a starting point for the theoretical evaluation of the behaviour shown in Fig. IV.7.

2.5.4 Applying the linear method to weakly non-linear GWPs

Our method relies on fitting $\partial_x c_{\mathrm{sp}}|_{x=0}$ as measured with crest and trough velocities to the simple profile of $\partial_x c_{\mathrm{sp}}|_{x=0}$ from Eq.(IV.6). In Sect. 2.5.3 showed that this profile is modified by higher-order non-linear effects. To apply our method, we also need to be able to measure $\partial_x c_{\mathrm{sp}}|_{x=0}$ based on crest velocity measurements, which can become impossible for very high values of $k_0 A_{\mathrm{f}}^{\mathrm{NL}}$. However, our method should still be applicable to weakly non-linear GWPs with a small enough amplitude. As we have seen in the previous section, higher-order non-linear effects on crest velocities seem to be small enough that our method is likely to provide consistent results when amplitudes are below the approximate threshold $k_0 A_{\mathrm{f}}^{\mathrm{NL}} \leq 0.16$.

Following this reasoning, we applied our method directly to estimate t_{f}, $A_{\mathrm{f}}^{\mathrm{NL}}/A^{\mathrm{NL}}(t)$ and $k_0 L_{\mathrm{f}}^{\mathrm{NL}}$ over non-linear wave packets with parameters $k_0 L_{\mathrm{f}}^{\mathrm{L}} = 5$ and $k_0 A_{\mathrm{f}}^{\mathrm{L}} = 0.13$

(which correspond to $k_0 A_f^{\mathrm{NL}} = 0.16$ in our simulations of the NLSE). We repeated this estimation procedure over several wave packets with different values of the constant and uniform phase (40 values seemed enough to capture the variability of the results). We added only a minor change to our method. As in the previous section, we measured the specular velocity gradient only when a crest or trough was lying close enough to the center ($|x| \leq 4\lambda_0 \mathrm{d}t/(2T_p)$). This provides fewer points for the estimation of the profile of $\partial_x c_{\mathrm{sp}}|_{x=0}$, but these points are of better quality, especially since the oscillations witnessed in Fig. IV.4 are strengthened by higher-order non-linear effects.

The results of this test can be seen in Fig. IV.8. For values of $t_{\mathrm{end}} \leq -23T_p$, despite some variability in the results, most points converge to a satisfying estimation of t_{f} and very large overestimation of $A_f^{\mathrm{NL}}/A^{\mathrm{NL}}(t_{\mathrm{end}})$ and underestimation of $k_0 L_f^{\mathrm{NL}}$. This inaccuracy is much larger than the errors witnessed in Fig. IV.5. This shows that more points are needed for the linear method to provide acceptable results for weakly non-linear wave packets than are needed for linear wave packets. For $t_{\mathrm{end}} \geq -23T_p$, the method converges to a fairly good estimation of $A_f^{\mathrm{NL}}/A^{\mathrm{NL}}(t_{\mathrm{end}})$ and $k_0 L_f^{\mathrm{NL}}$, while t_{f} is underestimated by a few wave periods. These results indicate a certain stability of our method with respect to higher-order non-linear effects, such that consistent results can still be achieved for weakly non-linear wave packets if enough measurements of crest/trough velocities are provided.

However, as we have seen, our method is not applicable for higher-order non-linear wave packets above a certain amplitude threshold. This threshold depends on $k_0 L_f^{\mathrm{L}}$, and our simulations show that for $k_0 L_f^{\mathrm{L}} = 5$ it is around $k_0 A_f^{\mathrm{NL}} \approx 0.16$. Precise determination of this threshold requires a more thorough examination. Note that the one-dimensional NLSE overestimates the increase in amplitude due to non-linear effects compared to more precise, higher-order evolution equations (Cousins and Sapsis, 2015a). Therefore, one might underestimate the threshold above which our method is not valid when using numerical simulations of the NLSE.

Above this threshold, generalization of our method is delicate and demands further study. A natural path would be to use analytical solutions for weakly non-linear GWPs such as those in Pizzo and Melville (2019). Another possibility would be to use a data-driven approach to learn how our method should be modified to take into account higher-order non-linear effects.

To investigate the consequences of such a threshold for real applications of our work in the ocean, we used the data analysed in Ruju et al. (accepted) for several strong storms, with H_s values ranging from 5 m to 20 m. With this data, we used the definition of design
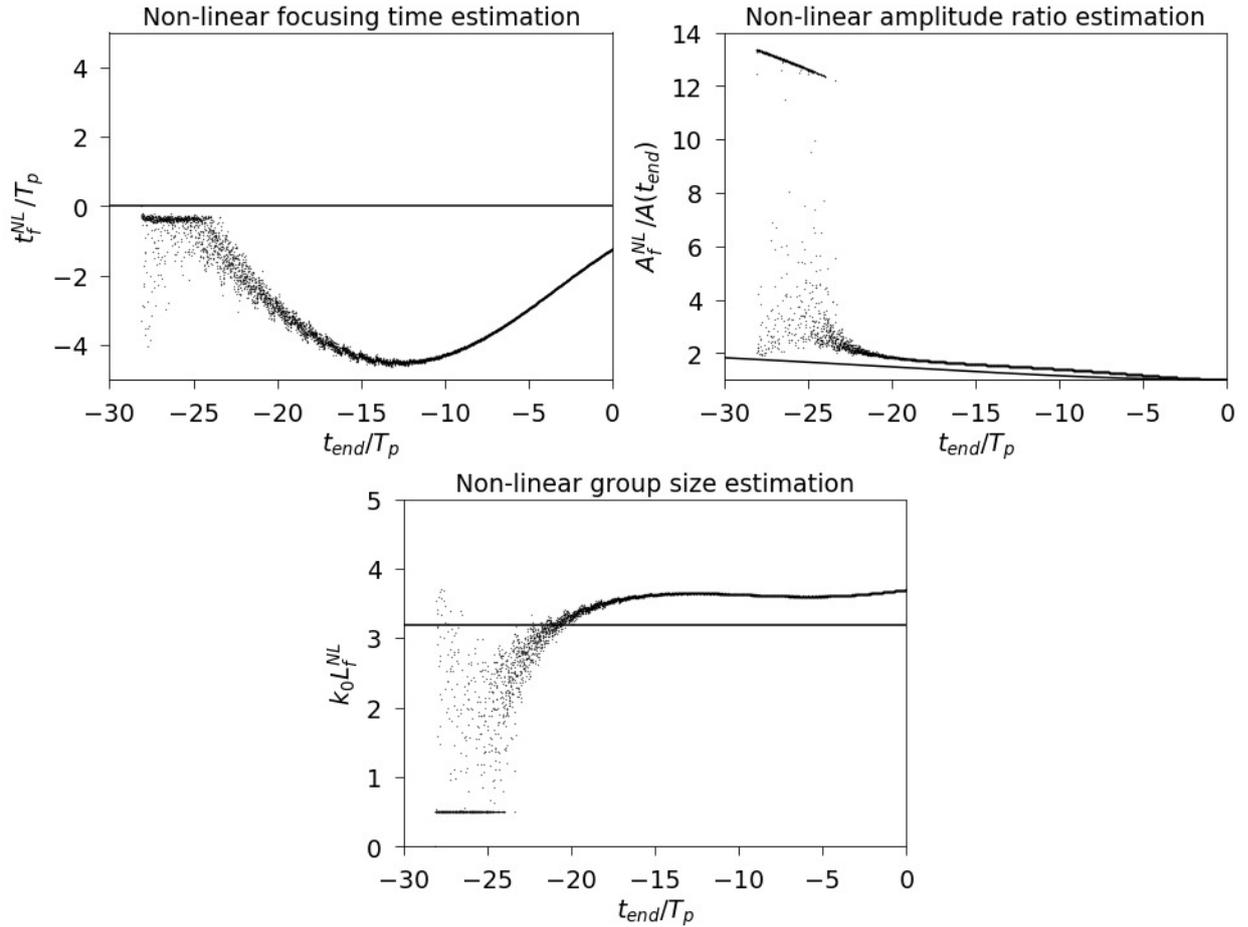
Figure IV.8: Estimation of t_{f}, $A_{\mathrm{f}}^{\mathrm{NL}}/A^{\mathrm{NL}}(t)$ and $k_0 L_{\mathrm{f}}^{\mathrm{NL}}$ for non-linear wave packets with parameters $k_0 A_{\mathrm{f}}^{\mathrm{L}} = 0.13$ and $k_0 L_{\mathrm{f}}^{\mathrm{L}} = 5$ ($k_0 L_{\mathrm{f}}^{\mathrm{NL}} \approx 3.2$). We use directly the method that was designed for linear wave packets and we start observing crest velocities at $t_{\mathrm{start}} \approx 30 T_p$. True values (objectives) are in full black lines. Estimations are in black dots. Here we present many estimations coming from simulations of GWPs with different values of the constant phase. The non-linear evolution of one-dimensional GWPs is approximated through numerical simulations of the NLSE.

wave used for offshore structures. This design wave is the largest wave encountered out of 1000 waves, on average, and equals $1.86 H_s$ for linear waves. While maintaining awareness of the limitations of such an approximation, we use it to find an order of magnitude of $k_0 A_{\max}$, the largest adimensional wave amplitude one would be likely to face in such storms. By doing so, we found values of $k_0 A_{\max}$ mostly between 0.1 and 0.22, with a peak at 0.25. If the aforementioned threshold amounts to about 0.16, these values of $k_0 A_{\max}$ indicate that our method could be used for many waves encountered in these storms, but it would need some modifications to be applicable to the steepest storm waves.

2.6 Conclusion

We have shown that the focusing time, the ratio of amplitude at focus to instantaneous amplitude, and the width at focus of an isolated GWP can be inferred from the measurement of crest and trough velocities over time near the center of the GWP. Our analytical derivations assume the linear dispersion of one-dimensional ocean surface gravity waves. The proposed methodology has been successfully tested over numerically simulated linear Gaussian wave packets with spectral widths of 0.15, 0.071 and 0.035, therefore proving relevant in terms of spectral width for dominant waves in storm-like conditions.

The linear approximation breaks down for real surface waves of finite amplitude. However, this work is still directly applicable to second-order non-linear wave packets. Analysing the crest/trough velocities of higher-order non-linear GWPs simulated numerically with the NLSE, we found that our method shows a certain stability to higher-order non-linear effects, and that it becomes inapplicable above a certain amplitude threshold. The evaluation of this threshold was not performed here and would require a thorough investigation, as it depends on the spectral width of the GWP. Above this threshold, our method could be generalized using either approximate analytical models for the evolution of higher-order non-linear wave packets, or data-driven tools from the machine learning community.

Still, in many situations the forecasting of rogue waves in moderate seas may be helpful. This is the case for marine operations at sea or transfer of workers from a ship to an offshore wind turbine.Our decision to rely solely on crest/trough velocities was motivated by both the sea surface reconstruction technique using lidar illumination and the combination of X-band radar images and stereo video cameras.

This study is a starting point to further investigate the applicability of our scheme to realistic random ocean surface gravity wave fields. First, the interactions between

different wave packets were not considered here. Such interactions typically happen on larger time scales than wave packet contraction, and are therefore often put on the back burner. Secondly, the corrections to be added when taking directionality into account are an important concern. Directionality affects the higher-order non-linear changes to wave group contraction, mitigating the amplitude increase and inducing strong changes in the shape of the wave packet at focus (Adcock et al., 2012). Broadening the range of application of our approach is vital, as the combination of wave group interactions and directionality could be responsible for the formation of one of the largest waves measured at sea (Adcock et al., 2011).

The measurement of crest/trough velocities in short-crested seas is also obviously a technical challenge. However, to use our method on real wave fields from measurements, one could filter out the high frequencies that contain both instrumental noise, high frequency wind waves and second-order non-linearities. Since none of those are either responsible for or interacting with the long-wave crest velocity behaviour described in this study, the methodology should be applicable to the filtered measurements. Note also that wavelength could be used in the same way as crest velocities, as wavelength is proportional to the square of the velocity for long-crested waves, and wavelength can be measured approximately from zero down- and up-crossings, which is less sensitive to high-frequency noise. In this case, the analog of the gradient of crest velocity would be the gradient of wavelength. This suggested counterpart methodology could be investigated in further research, but is yet likely to yield other unexpected limitations.

Wave energy dissipation and wind energy input are usually considered negligible for the forecast horizons considered here, which are inferior to approximately $60\,T_p$. In fact, integrated parameters of the wave spectrum such as H_s are usually estimated from data collected during approximately $100\,T_p$ (e.g. Mitsuyasu et al., 1975). Forecast above this threshold of $\approx 60\,T_p$ is not considered for two reasons. First, instrumental noise makes it hard to forecast individual waves more than $100\,T_p$ ahead (see the predictability horizon introduced by Alam, 2014). Second, it is hard to imagine measuring waves more than 60 peak wavelength ($\approx$10 km) away from a ship or an offshore platform using on-board measurement devices. For now, the spatial horizon of X-band marine radars is limited to $\approx$2.5km and is even smaller for stereo-video cameras (Benetazzo et al., 2018).

Numerical and wave-tank experiments of Kharif et al. (2008) showed that for large amplitude waves, wind-wave interactions can both weakly increase the maximum wave amplitude and noticeably shift the focusing point downstream. This further influence of

wind would have to be parameterized in addition to the proposed methodology.

Data accessibility

All the data and python scripts used to produce the figures of this article are publicly available in the following GitHub repository:

https://github.com/PaulPlatzer/Wave_Focusing_Crest_Velocities

3 Summary

☐ We characterized the focusing or defocusing nature of narrow-banded, Gaussian, linear ocean wave packets through crest and trough velocities. A focusing wave packet has faster crests at its front (downstream) and slower crests at its back (upstream).

☐ These spatial variations of crest velocity evolve in time according to a profile of which we give an approximate analytical expression. Observations of crest velocities can then be fitted to this profile to estimate the time, location, and amplitude of the wave packet at focus.

☐ Numerical experiments of linear Gaussian wave packets confirm the applicability of our methodology. Observing crest velocities during a few wave periods allows us to estimate the focusing point, time, and relative amplitude within acceptable confidence intervals.

☐ We provided analytical arguments for the applicability of our method to second-order non-linear Gaussian wave packets.

☐ Numerical experiments of the non-linear Schrödinger equation show a certain robustness of our method to higher-order non-linear effects. For weakly non-linear Gaussian wave packets, our methodology is still applicable, although a larger observation time-window is necessary to obtain acceptable estimations, and we still witness higher errors than with linear Gaussian wave packets.

☐ For wave packets with steepness $k_0 A \gtrsim 0.16$, our methodology is no longer applicable, and higher-order non-linear effects must be accounted for. Nonetheless, our methodology seems to be applicable to a large range of real-world sea-states.

The influence of wave energy intake and dissipation, local wind, surface currents, wave group interactions, and directionality was left for further research. A more thorough study of higher-order non-linear wave effects is necessary. Finally, using a statistical framework to associate confidence intervals to our estimations would be interesting, as our objective is to assess the risk of encountering an extreme wave. Both observation and model error would have to be accounted for.

The methodology described in this chapter is based on physical knowledge of the motion of the ocean surface. One could also imagine to try to forecast extreme wave, wind or rain events using statistical tools such as the one presented in Chapter III. However, using empirical, data-based techniques to forecast extreme events requires to take some precautions, as empirical methods might be biased in the case of extreme events. For instance, it is natural to believe that using historical data to forecast future states might induce a bias towards mean states, which are more represented than extreme events. Also, when concerned with the forecast of extreme events, one must focus on distribution tails rather then mean and variance. The next chapter examines the use analog forecasting strategies for extreme events, gives limitations to the results of Chapter II and III, and paves the way for future research.

Can analogs forecast extreme events?

As now can't reveal the mystery of tomorrow
But in passing will grow older everyday
Just as all is born is new
Do know what I say is true...

– Stevie Wonder, *As*

1 Introduction

Heat waves, hurricanes, rogue waves... Extreme events of geophysical systems can have dramatic consequences, and the motivation for their forecast is obvious. Forecast strategies that are efficient on a regular basis may not succeed in the case of extreme events. In some cases, one can use special knowledge of the system's dynamics, such as a physical description, to forecast especially those extreme events. This was the strategy presented in Chapter IV. If one is not provided with such a physical model, one may use data-driven methods such as analog forecasting. Still, mathematical concepts and assumptions are needed to infer predictive probability distributions of extremal behaviours outside of the range of observed time series. Special probability theory tools can help to model extreme events.

Here, we consider heavy-tailed random variables, for which the probability to observe an extreme event is higher than for light-tailed variables. For example, hourly heavy rainfall are not normally distributed, and the probability of rare events decreases much slower than exponential decay. We investigate the following questions:

- Analog forecasting has proven efficient for Gaussian-like random variables. Can it be applied as it is for heavy-tailed random variables?

- The works of Van Den Dool (1994), Nicolis (1998), and of the second chapter of this thesis relate the typical analog-to-target distance, the size of the catalog (or database), and the dimension of the system. Do these tools apply to heavy-tailed random variables?

To answer these questions, the following strategies and tools are used:

* Trajectories of heavy-tailed random variables are generated from a one-dimensional state-space model. Particle filters allow to estimate the future state and reconstruct the hidden state. Forecasts from the model equations are compared with analog forecasts using various catalog sizes.

* The numerical experiments of Chapter II are repeated for numerically-generated samples of a one-dimensional heavy-tailed random variable.

2 Analog forecasting of heavy-tailed random variables

2.1 Context

There are several ways of defining heavy-tailed probability distributions. Both rigorous and intuitive definitions can be found in the book of Resnick (2007). For our purpose, it is enough to say that X is a heavy (right) tailed, positive random variable if its k-th moment is infinite for some $k > 0$. Heavy-tailed probability distributions include the inverse-gamma, Cauchy, Fréchet and Weibull distributions. These last three distributions arise naturally when considering threshold exceedances of random variables. Also, some physical quantities of interest have heavy-tailed probability distributions, such as rainfall in south of France (Taillardat et al., 2019). A detailed Bayesian treatment of rainfall extremes with a hierarchical spatio-temporal model can also be found in Bacro et al. (2020). These authors did mainly focus on parameters estimation, but they did neither investigate forecasting issues nor implement analogs techniques.

The third chapter gave a particular focus on first and second moments of analog forecasts. The underlined key assumption was the Gaussian nature of the studied quantities. In this chapter, we move away from normality and consequently, from first and second moments to the full analogs distributions. Still, analog forecasting operators may be able

to handle skewed distributions or multi-modality if they are viewed as empirical distributions, i.e. Dirac sums. The next section presents numerical experiments where analog forecasting operators are used to make predictions of heavy-tailed random variables.

2.2 One-dimensional toy model

To our knowledge, bridging extreme value theory and analog approaches has never be done for heavy tailed distributions. In this context, our goal is to study univariate stationary time series based on "simple" dynamical structures that insure Pareto-distributed observations. To go further, e.g. studying extreme compound events (see Zscheischler et al., 2020), one will need to couple multivariate extreme value theory and analog techniques, a challenging task left for future work.

We consider the following one-dimensional state-space model. Observations Y_t are taken from a latent state X_t. The time t takes integer values, and X_t and Y_t are real-valued one-dimensional random variables.

$$X_t = \mathcal{M}_t(X_{t-1}) \,, \tag{V.1a}$$

$$Y_t = \sigma \frac{E_t}{X_t} \,, \tag{V.1b}$$

where $\mathcal{M}_t$ is a random mapping from X_{t-1} to X_t that takes different shapes. In all the applications presented hereinafter, the observation equation is assumed to be known. For all the models $\mathcal{M}_t$ considered, the margins of X_t are gamma-distributed, $X_t \sim \text{Ga}(\xi^{-1}, \xi^{-1})$ with real-valued parameter ξ. Also, X_t has autocorrelation in time $\text{Corr}(X_t, X_s) = \rho^{|s-t|}$, with real-valued parameter ρ. All the models $\mathcal{M}$ used here are taken from the work of Wolpert (2016).

From these latent gamma processes are drawn log-linear observations Y_t. The latter is obtained by drawing an independent unit exponential draw, $E_t \sim \exp(1)$, divided by the gamma process. The real-valued parameter σ controls the noise amplitude. The observation equation (V.1b) implies that Y_t follows a generalized Pareto distribution $Y_t \sim \text{GPD}(\sigma, \xi)$. Large observation values happen for small values of the latent space variable.

Latent AR1 gamma process :

$$\mathcal{M}_t^{\text{AR1}}(X_{t-1}) = \rho \, X_{t-1} + \eta_t \,, \tag{V.2}$$

157

where η_t is iid and has characteristic function $u \mapsto \mathbb{E}e^{iu\eta_t} = \left[\frac{1-u\xi}{1-\rho u\xi}\right]^\xi$. This model belongs to the class of first-order auto-regressive models.

Latent thinned gamma process :

$$\mathcal{M}_t^{\mathrm{THIN}}(\mathrm{X}_{t-1}) = \mathrm{B}_t\,\mathrm{X}_{t-1} + \zeta_t\,, \tag{V.3}$$

where B_t and ζ_t are i.i.d. and follow respectively a Beta distribution $\mathrm{B}_t \sim \mathrm{Be}(\rho/\xi, \overline{\rho}/\xi)$ and a gamma distribution $\zeta_t \sim \mathrm{Ga}(\overline{\rho}/\xi, 1/\xi)$ with $\overline{\rho} = 1 - \rho$ (Wolpert, 2016).

Latent squared Ornstein-Uhlenbeck gamma diffusion :

$$\mathcal{M}_t^{\mathrm{OU}}(\mathrm{X}_{t-1}) = \mathrm{X}_{t-1} + \lambda \int_{t-1}^t (\mathrm{X}_s - 1)\mathrm{d}s + \sqrt{2\lambda\xi} \int_{t-1}^t \sqrt{\mathrm{X}_s}\mathrm{dW}_s\,, \tag{V.4}$$

where integrals are taken in the Itô sense and dW is a Wiener process or Brownian motion. Here $\lambda = -\ln\rho$ and we have corrected for a forgotten '1/2' in Wolpert (2016). Such integrals correspond to stochastic differential equations that have unique strong solutions and can be solved numerically using, for instance, a Milstein method.

Example time series and $(t, t+1)$ relationships from these models are shown in Fig. V.1. It can be seen that although Y_t and Y_{t+1} appear to be independent, there is a clear relationship between X_t and X_{t+1}, with a different behaviour for each model.

From these models, one would like to estimate the following probability distributions:

$$\mathrm{d}\mathbb{P}(\mathrm{X}_{1:t} = x_{1:t}\,|\,\mathrm{Y}_{1:t} = y_{1:t})\,, \tag{V.5}$$

which is the probability density function of latent space reconstruction from past observations (from time 1 to t), and:

$$\mathrm{d}\mathbb{P}(\mathrm{Y}_{t+1} = y_{t+1}\,|\,\mathrm{Y}_{1:t} = y_{1:t})\,, \tag{V.6}$$

which is the probability density function for the forecast of the next observation knowing all past observations. Other types of quantities could be estimated, such as the probability of exceedance of a threshold for the next observation. However, for a first investigation, it was chosen to focus on full probability density functions.
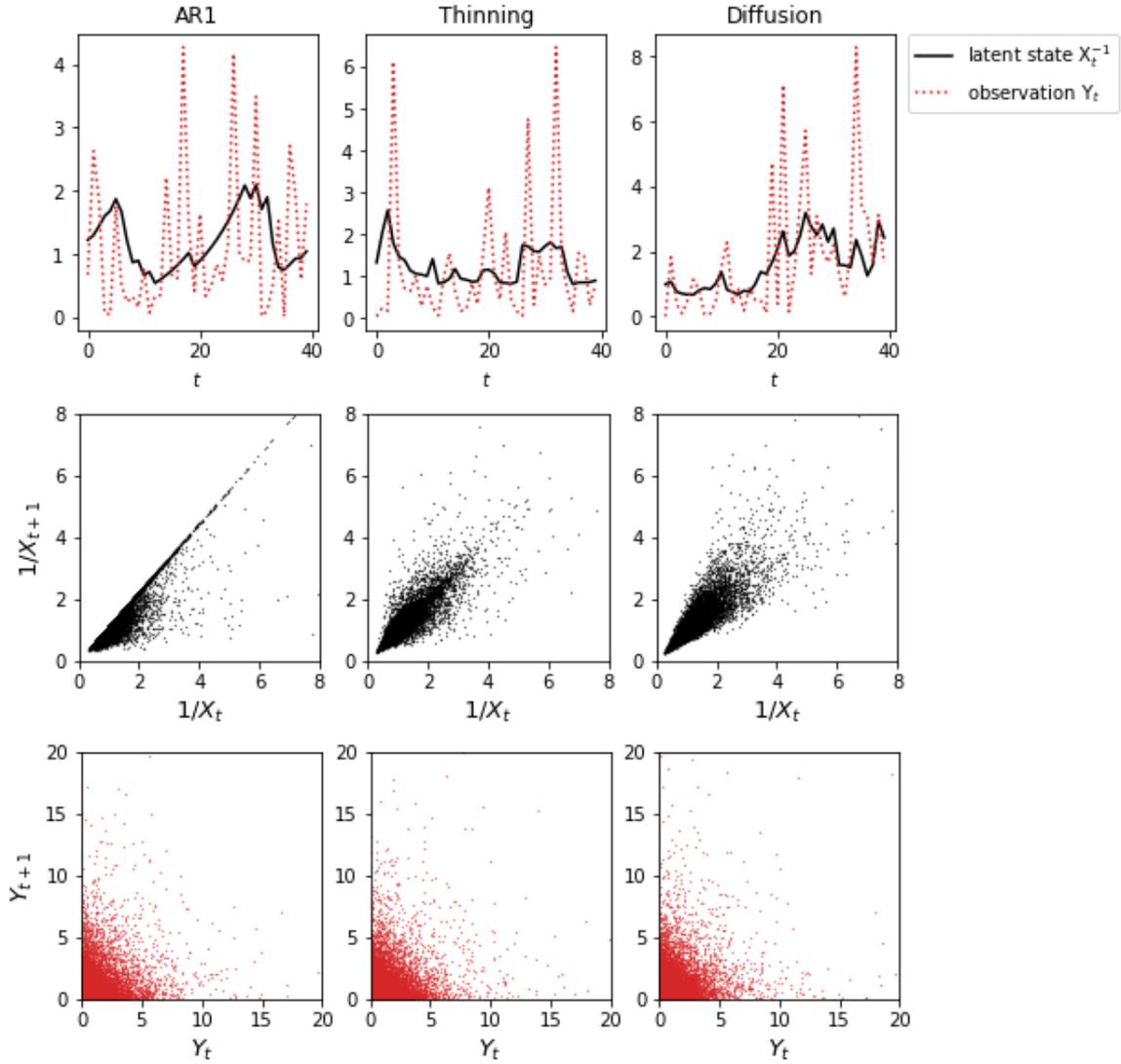
Figure V.1: Top panels: example time series generated from the models (V.2), (V.3) and (V.4) and observation (V.1b). Middle panels: $(t, t+1)$ relationship in latent space. Bottom panels: $(t, t+1)$ relationship in observation space. Scale parameter $\sigma = 1$, shape (tail) parameter $\xi = 0.2$, and correlation $\rho = 0.9$.

2.3   Crude analog forecasting with particle filters

One way to estimate the probability distributions (V.5) and (V.6) is to use data assimilation. As the system at hand clearly does not produce normally distributed solutions, the classical Kalman filter equations cannot be applied. Different techniques to handle non-Gaussian systems have been proposed (see, e.g., book by Cappé et al., 2006). Particle filters (Van Leeuwen, 2009) are able to reconstruct probability distributions of any shape, provided that the number of particles is large enough. Assuming that one knows the model equation in latent space (V.1a) and the observation equation (V.1b), particle filters allow to estimate the quantity (V.5), and from samples of the probability distribution of X_t one can draw samples of Y_{t+1} using the model state and observational equations.

This classical particle filter strategy can be compared with the same forecast scheme where the model $\mathcal{M}_t$ is replaced by analog forecasting. If one is provided with samples of latent space variables, i.e. a catalog of latent space $\mathcal{C} = \{x_i, i \in [1, L]\}$ which follow the random mapping $\mathcal{M}_t$, then the model equations (V.1a) can be replaced by analog forecasting operators presented in the third chapter, using the catalog $\mathcal{C}$. One might argue that assuming to have a catalog of latent-space variables rather than observations is a strong and somewhat irrelevant hypothesis, but it is taken here as a first step. The reader is further referred to Chau et al. (2020) for applications of analog forecasting strategies using a catalog in observation space rather than latent space. In a lighted-tail context, the authors were able to reconstruct a catalog of latent-space trajectories using data assimilation schemes, and this catalog of reconstructed latent-space trajectories was then used to emulate $\mathcal{M}_t$.

In our example, particle filters with 100 particles are used to reconstruct the probability distribution of the latent space variable (V.5) and of the next observation (V.6), for a sequence of latent space states and observations simulated from the AR1 model (V.2) and the observation equation (V.1b) with parameters $\xi = 0.2$ (moderately heavy-tailed, similar to rainfall variables, Naveau et al. 2016), $\sigma = 1$, and $\rho = 0.9$. Recall that the observation equation is assumed to be known in all cases. The time-evolution in latent space is either simulated directly from the "true model" (V.2), or from "misspecified model 1" (V.3) or "misspecified model 2" (V.4), or from analogs, using the locally-constant operator presented in Chapter III. Each analog forecast is performed using 40 analogs, drawn either from a "large catalog" of size $L = 10^5$, a "medium catalog" of size $L = 10^3$, and a "small catalog" of size $L = 40$, all simulated from (V.2).

Example forecasts are shown in Fig. V.2. The particle filters with model equations and

with analogs (large and medium catalogs) seem to be able to trace some of the tendencies in the observation given by the value of X_t, with quite similar forecast quantiles, both in latent and observation space. However, due to the multiplicative observational noises, the particle filters struggle to capture the real state. The analog particle filtering with a very small catalog (only 40 possible analogs) is unable to reproduce the tendencies of X_t. Since every analog is used at each forecast ($K = L = 40$), the variability of the latent space is overestimated. For these medium values of $1/X_t$, it is hard to predict the future observation Y_t, and all forecasts are quite similar.

To compare the efficiency of analog forecasting with model-based forecasting, one can use the following score:

$$s\left(d\tilde{\mathbb{P}}\right) = \frac{1}{t} \sum_{k=1}^{t} \log d\tilde{\mathbb{P}}(y_{k+1}), \qquad (V.7)$$

where $d\tilde{\mathbb{P}}$ is the estimated probability density functions implicitly conditional on the past observations $(y_1, \ldots, y_k)$. A low score corresponds to a high log-likelihood and is therefore better than a high score. For large t, the difference score between two different estimates $d\tilde{\mathbb{P}}$ converges to the Kullbak-Leibler divergence between the probability density functions of these estimates.

Fig V.3 shows the behaviour for large t of the difference in forecast score between the particle filters using either misspecified models or analogs and the reference particle filter using the true model equations. All scores are close to ≈ 1.6 (not shown), such that the values of the Kullback-Leibler divergences are quite small, meaning that the scores are quite similar, using either the true or misspecified model equations, or large or small catalogs. For small values of t, one of the misspecified models gives better scores than the good model, and the analog method with medium catalog gives better results than with a large catalog. However, this changes around $t \approx 20000$ and again around $t \approx 95000$, where all scores jump.

These jumps are the result of very large values of Y_t (due to large values of $1/X_t$), as can be witnessed in Fig. V.4. For such large values, only the true model equations or the large catalog are able to trace the system state. Note that $1/X_t$ is slightly underestimated, which might be a consequence of the fact that the particle filters track the state $1/X_t$ and not its inverse. Performing particle filters using another representation, with state-space $1/X_t$ and observations $\log(Y_t)$, might give different results.

It is worth noting that the simulation had to be performed for a very long time before
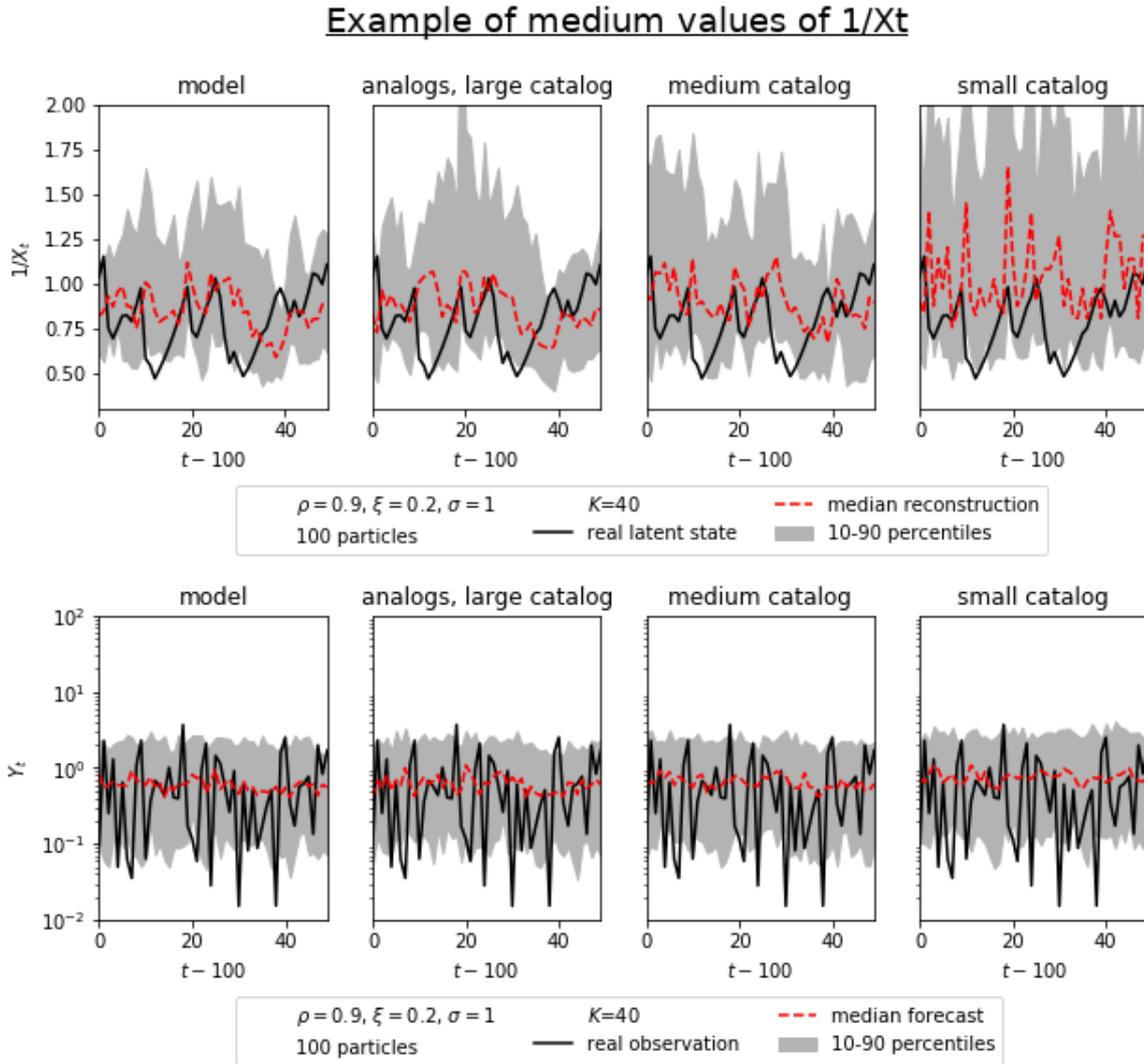
Figure V.2: Applying the particle filter with 100 particles to reconstruct the latent space and forecast the next observation. Comparison of the filtering with known model Eq. (V.1a), or when this equation is replaced by the locally-constant analog forecasting operator in latent space. Catalog sizes are $L = 10^5$, 10^3, 40. Upper panels: $1/X_t$. Lower panels: Y_t.
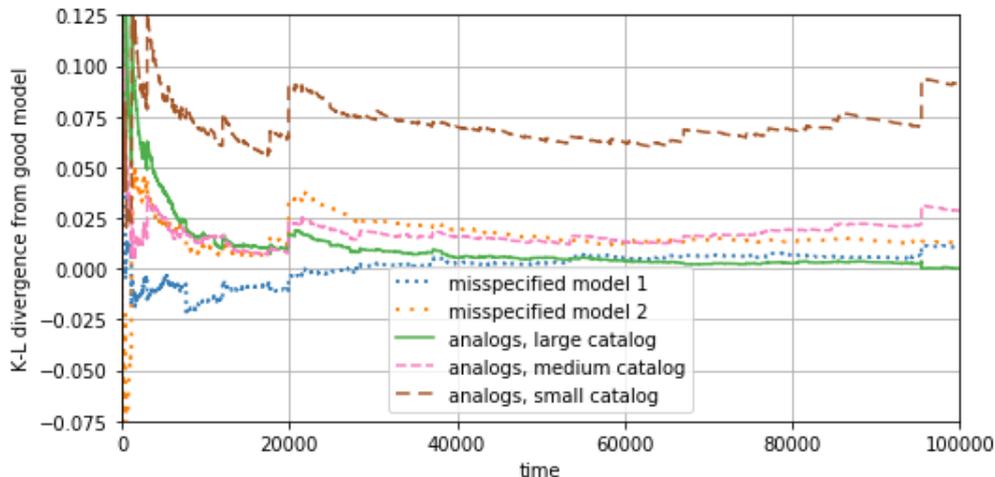
Figure V.3: Difference of forecast score between several methods, all using particle filters, and the reference classical particle filtering technique using the real model equation (V.2), averaged over t forecasts.

reaching such a high value of $1/X_t$, and that only then did the true model equations or the large catalog reveal their usefulness in estimating the next observation. Also, this experiment shows that these very high observations completely change the average score of each method, although they are very rare. This behaviour is typical of heavy-tailed random variables. It shows that one can design a method that seems efficient for "normal" events and eventually fails when faced with an "extreme" event.

For the sake of brevity, only experiments using the AR1 model were shown, but one could also have used the other models presented here which have different temporal behaviours, and our methodology is not restricted to this very specific AR1 case. Also, one could have used locally-incremental or locally-linear analog forecasting operators. Numerical experiments (not shown) indicate that these methods allow to increase the variability of the forecasts compared to the locally-constant, but the forecasts are somewhat inconsistent and do not seem to emulate well the models $\mathcal{M}_t$, such that the locally-constant was the preferred choice in this case.

2.4 Analog-to-target distances

In Chapter II, the theoretical distribution of analog-to-target distances was given. This theoretical result was based on dynamical systems theory and extreme value theory ap-
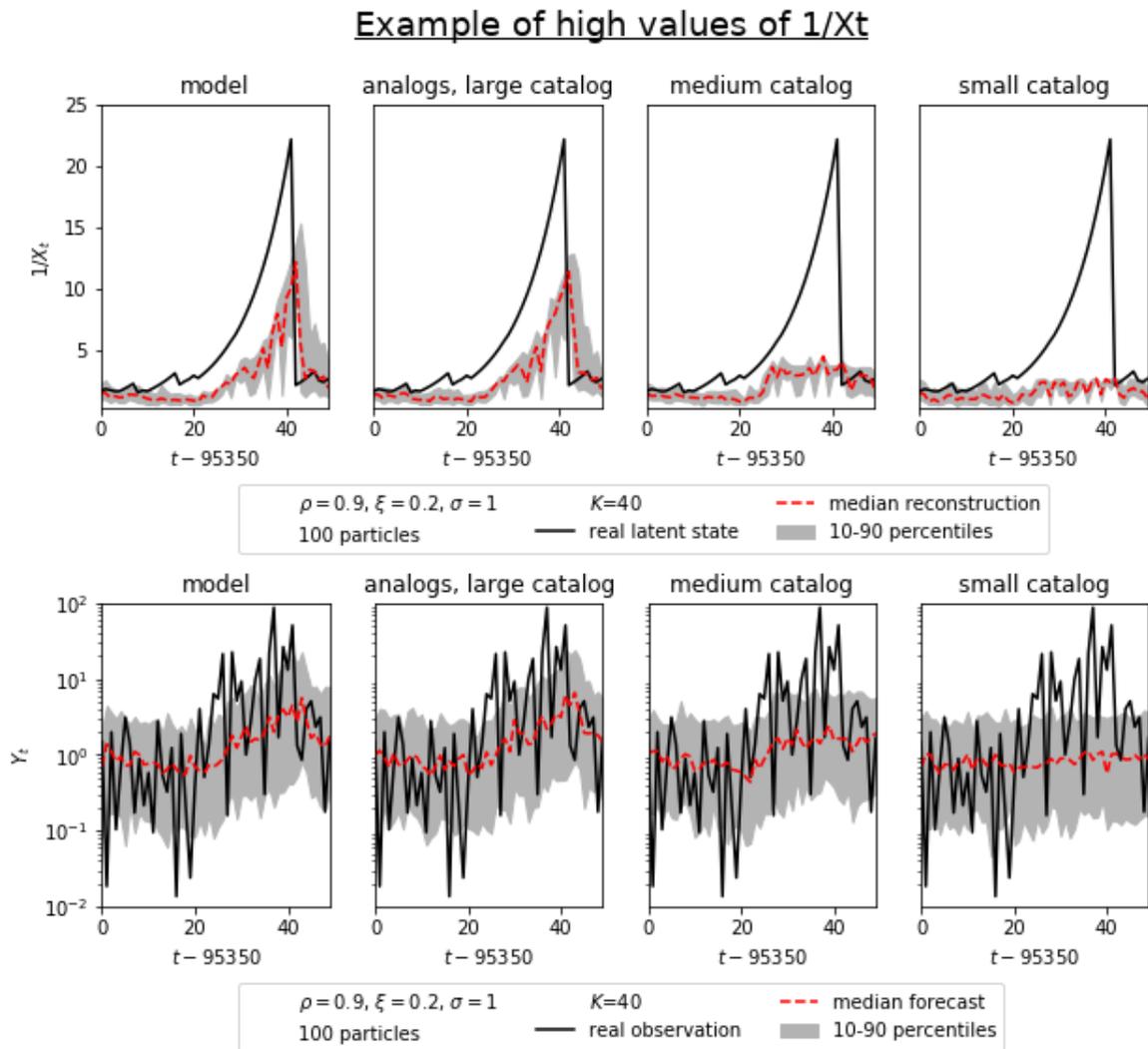
Figure V.4: Same as Fig. V.2 but for large values of the inverse latent state $1/X_t$ that cause average scores of different methods to jump (see Fig. V.3 around $t \approx 95000$).

plied to a specific distance observable. However, our result does not seem to be applicable to the case where the target state is an extreme event.

In particular, we consider the inverse of the state-space variable $1/X_t$ introduced in the previous section. As X_t is gamma-distributed with parameters $\alpha = \beta = 5$, $1/X_t$ follows the inverse-gamma distribution with same parameters, and is thus heavy-tailed. Then, we search for analogs of a given value of $1/X_t$, inside a catalog generated from the same process.

If the target value is not too high, and one applies the methodology of Fig. II.3 in Chapter II, one finds similar results, with estimated dimensions around ~ 1, as shown in Fig. V.5. The estimation of dimension is independent of the catalog size, and so is the scaling factor ρ (not to be confused with the time-correlation introduced in this chapter). The scaling of $C \sim L^{-1/d}$ is recovered, and the distributions from Eq. (II.11) seem to apply, even though not perfectly for the case of $k = 30$ and $L = 10^2$, but this can be explained by the fact that the limit of large catalog is hardly achieved in this case.

Then, one can try to apply this methodology to an extreme event, i.e. a large value of $1/X_t$, as shown in Fig. V.6. In this case, the estimation of dimension strongly depends on catalog size, and so does the scaling factor ρ. For large catalog size $L = 10^4$, the estimated dimension is close to the value of 1, which is consistent with the fact that we are searching for analogs of a one-dimensional variable. However, estimated values of the scaling factor are of the order of 200, which indicates that the scaling $C \sim L^{-1/d}$ is not recovered. The empirical probability distribution of 1st, 10th and 30th analog-to-target distances seem to fit the theoretical distributions for $d = 1.2$, which is the mean of estimated dimensions. For lower values of the catalog size $L = 10^3$, 10^2, the estimated dimensions are different from the real dimension (=1), such that there is an "apparent" higher dimension around such extreme events. The scaling factor exhibits lower values, indicating that the scaling $C \sim L^{-1/d}$ is valid if one uses this "apparent" dimension. The empirical probability distribution of 1st, 10th and 30th analog-to-target distances show behaviours similar to the theoretical distributions with mean values of estimated dimension and scaling factor, although this would not be expected since this example is one-dimensional.

This small experiment clearly shows that the theory of Chapter II must be adapted for probability distributions of analog-to-extreme-event distances. As the scaling $C \propto L^{-1/d}$ seems not to be valid for values in the tail of heavy-tailed random variables, the adapted theory should be able to give the scaling of C with L, d, and possibly parameters of the tail of the distribution. This adapted theory could build on the work of Holland et al.

Figure V.5: Applying the methodology of Chapter II to heavy-tailed random variables. When searching analogs of values that are not in the tail of the distribution, the methodology seems to be applicable. In this figure, ρ is the scaling factor introduced in Chapter II, not the correlation parameter introduced in Chapter 5.
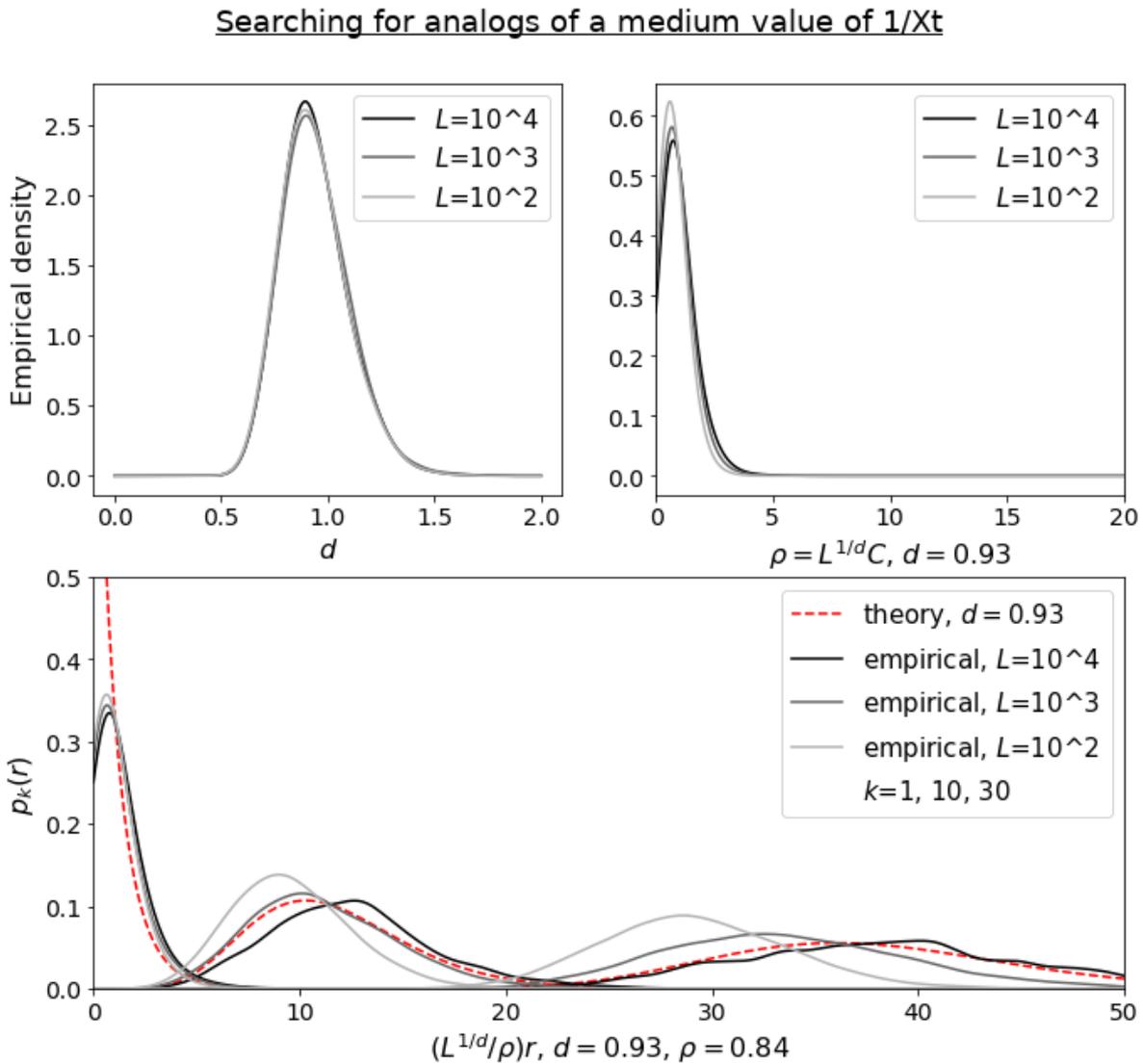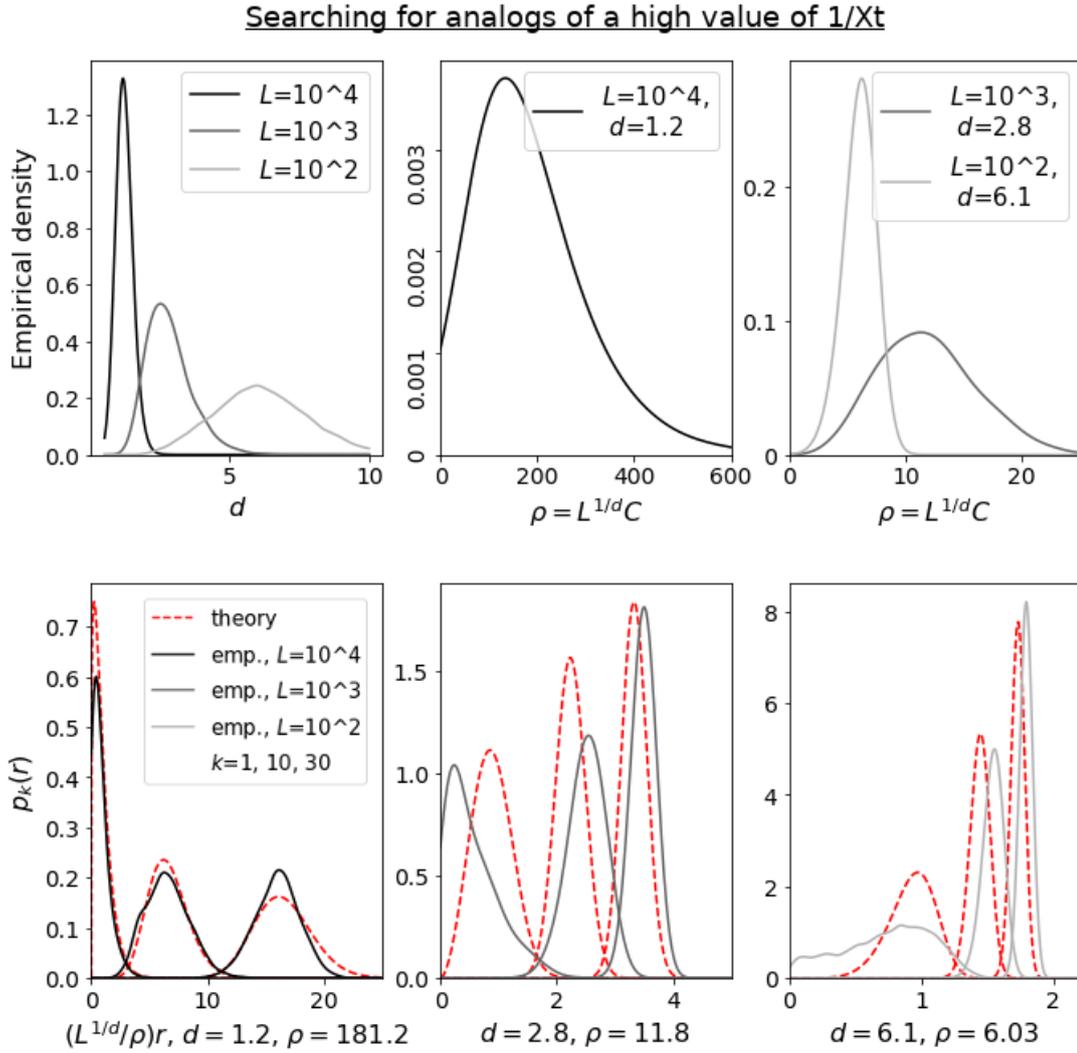
166

Figure V.6: Applying the methodology of Chapter II to heavy-tailed random variables. When searching analogs of values that are in the tail of the distribution, the methodology seems to fail. In this figure, ρ is the scaling factor introduced in Chapter II, not the correlation parameter introduced in Chapter 5.

(2012) and Lucarini et al. (2014) on extreme value theory for "physical" observables of dynamical systems. Also, the recent work of Caby et al. (2020) on the fractal structure of extreme values for observables of dynamical systems might help understanding the behaviour of "apparent" dimension reported in this last section.

2.5 Perspectives

The last section evaluated the performances of a *classical* analog forecasting operator, in its simple empirical form, for the forecast of a one-dimensional heavy-tailed random variable. Here, we draw perspectives on how analog forecasting strategies could be adapted (i.e. building "non-classical" analog forecasting tools) to better estimate the probability of extreme events.

The analog forecasting strategy used in the last section can be seen as a Monte-Carlo sampling technique. It is known that the Monte-Carlo estimation of an infinitesimally small probability demands an infinitely large number of samples (Morio et al., 2014), which corresponds here to an infinitely large number of analogs. How could the sampling efficiency of analog methods be improved for rare events?

In practice, one is provided with a limited catalog, such that one is *not able to resample*, i.e. to produce new analogs. However, the way the analogs are sampled from the (fixed) catalog is a matter of choice. Fig V.7 shows schematically, on a two-dimensional example, how classical sampling based on analog-to-target distance is suboptimal in the context of extreme events. For a more efficient sampling of analogs, one would like to select the analogs that contribute to the small probability of encountering an extreme event. Sampling techniques used to select weights in particle filtering could be tailored at extreme analogs (see e.g. Toulemonde et al., 2013). However, in this case, the tuning of the weights becomes very important. In the conventional analog forecasting strategies of Chapter III, the weights ω_k are usually simple decreasing functions of the analog-to-target distance, that allow to discard analogs that are too far away from the target state. Conversely, the probability of encountering an extreme event might be entirely determined by these analogs, so that the precise estimation of the low weights ω_k is crucial. It follows from importance sampling techniques (L'Ecuyer et al., 2009) that these weights should be equal to the original probability density function of the future state evaluated at the successor's position, divided by the number of analogs K. However, the original probability density function is generally unknown. This indicates that further research is needed to sample rare events from analogs.
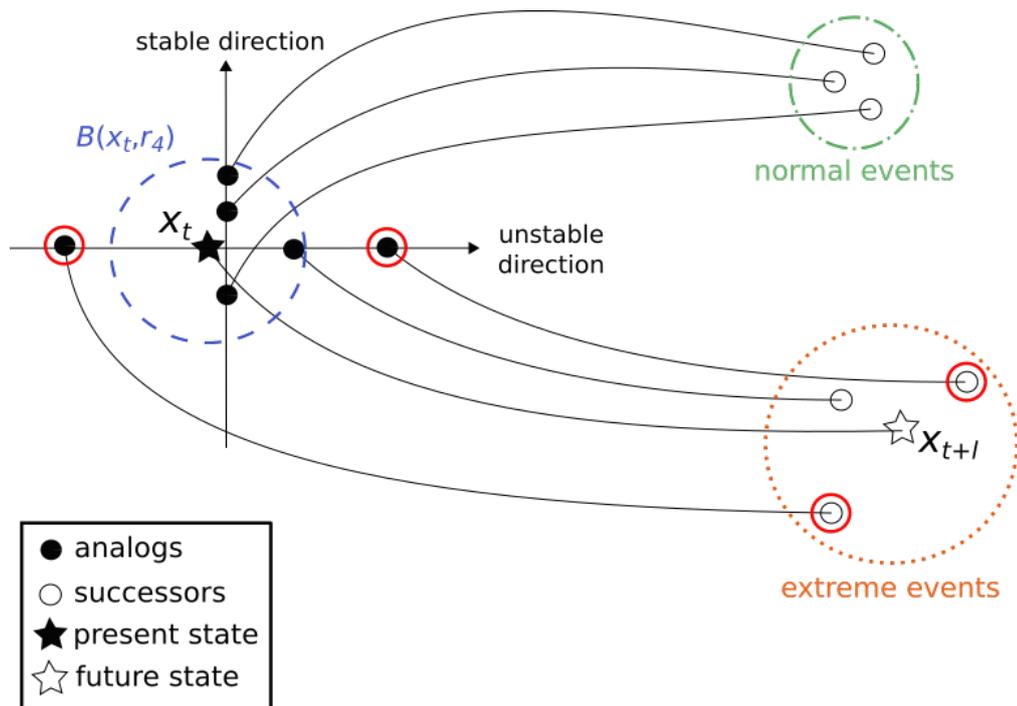
Figure V.7: A situation where classical sampling of analogs (with respect to distance) is suboptimal for the forecast of extremes. The two analogs circled in red are not taken into account if $K = 4$, while they would have provided relevant information on the probability of encountering an extreme event.

Since analog resampling is limited, one can also consider parameterization techniques. For instance, if the random variables are known to be Gaussian, then the parameters can be estimated as in Lguensat et al. (2017), which allows to compute the tail of the distribution. In a more general setting, the tail of the distribution could be estimated using results from extreme value theory to generalize forecasts outside the catalog of observed events.

3 Summary

☐ For heavy-tailed random variables with shape parameter $\xi = 0.2$, analog forecasting with a large catalog seems to be able to reproduce model-based forecasts. This can be attributed to the non-parametric nature of analog methods that allow to estimate whole probability distributions.

☐ However, both model-based and analogs-based particle filtering fail in generalizing above the highest observed state. Thus, more efforts are needed to couple extreme value theory and particle filters.

☐ When considering heavy-tailed random variables, analog forecasting with medium-sized and with large-sized catalogs give similar performances, except for some "outliers" (i.e. excursions of the system in the tail of the distribution). Despite the fact that these outliers are rare, they are the main driver of the difference in average score between analog forecasts with medium-sized and with large-sized catalogs. Raising the catalog size allows to better approach these outliers and thus gives a better averaged forecast score, although it gives similar forecasts most of the time. This suggests that, to forecat extreme events of heavy-tailed random variables using classical analog forecasting techniques, the catalog size must be increased beyond typical values for light-tailed random variables.

☐ For the heavy (right) tailed, one-dimensional random variables considered here, the probability distributions of Chapter II hold for moderate values of the random variable. If one then searches for analogs of higher values of the random variable, using the same catalog, the distributions no longer hold. In particular, the scaling of analog-to-target distance with dimension and catalog size is not valid. Furthermore, the estimated dimension is higher than the real space dimension. This suggests that the theory of Van Den Dool (1994), Nicolis (1998), and of the second chapter of this

thesis must be adapted to assess the probability of finding analogs of extreme events, taking into account tail parameters.

CONCLUSION AND PERSPECTIVES

Results summary

In this thesis, I contributed to the field of geophysical forecasting, using both data- and model-based methods, with a particular focus on analogs and large amplitude ocean waves.

In Chapter II, I gave an analytical expression for the probability distributions of analog-to-target distances. These allow to estimate the probability to find good analogs of a given state for a given dynamical system. The proposed method is based on the local dimension and the size of the catalog in which analogs are sought for. I showed how these probability distributions can be useful to any analog method. In particular, the variability of analog-to-target distances appears to be a decreasing function of the dimension. Consequently, in high dimensional problems the best analogs of a given point are "all as good as each other", while in low dimensions the first-best analog is likely to be much "better" than the, say, 30th-best analog. These results are a direct consequence of recent works combining dynamical systems and extreme value theory. Tests on numerical simulations of the three-dimensional Lorenz system and on wind maps from a physical model confirm the applicability of our analytical probability distributions to real geophysical data.

In Chapter III, I gave an interpretation of analog forecasting errors using the flow of the system. This interpretation aims at bridging data-based analog forecasting and model-based forecasts. Mean analog forecasting errors were shown to be linear in the analog-to-initial state distance, and quadratic for analog methods that make use of local linear regressions (i.e., using a locally-linear operator). It was shown that these linear regressions actually estimate the Jacobian matrix of the real system's flow map. I interpreted the locally-constant and locally-incremental operators as particular cases of the locally-linear operator, when the flow at a given time is either constant in phase-space, or equal to the identity flow. This explains the accuracy of the locally-incremental at small lead times. These analytical results were confirmed by numerical experiments of well-known chaotic dynamical systems. The robustness of such results to additive noise was also examined theoretically and numerically, justifying the use of a larger number of analogs to mitigate the influence of noise. The possible advantage of splitting a forecast into multiple analog

forecasts was lastly examined, showing promising preliminary results when analogs are combined with a linear regression.

In Chapter IV, I focused on large-amplitude ocean waves called "rogue waves" that are a threat to ships and offshore structures. A model-based method for the forecast of such waves was presented. I showed that the future highest waves could be forecasted using crest velocities rather than entire wave fields. This would give an advantage in terms of computation time, by focusing only on the largest waves. However, the estimation of crest velocities from present state-of-the-art wave measurement techniques is still a challenge. Our analytical derivations give profiles of crest velocities assuming linear ocean waves, narrow-banded wave spectra and Gaussian wave packets. These results were confirmed by numerical experiments of linear waves, and a method to forecast the position and amplitude of the focused wave packet was tested. Limitations due to non-linear wave interactions were examined in numerical simulations of the non-linear Schrödinger equation.

In Chapter V, I tackled the issue of extreme events forecast from purely empirical analog methods. A one-dimensional toy model with heavy-tailed random variables was built, on which analog methods were tested in combination with particle filtering techniques. For random variables with a moderately high tail index, analog forecasts reached the efficiency of model-based forecasts. The size of the catalog needed to find analogs of extreme events appeared to be higher than expected from the results of Chapter II. This calls for an adaptation of the theory of Chapter II for analog-to-extreme event distances, accounting for tail parameters of the distribution.

Perspectives

Analogs of extreme events

In Chapter V, I said that the theoretical description of Chapter II must be adapted to account for rare events. I have derived such an adaptation for one-dimensional random variables with continuous probability densities. In this case, the scale factor ρ of Chapter II depends on the value of the probability density function evaluated at the target state. For events in the tail of heavy-tailed random variables, this factor becomes large. Further work is needed to confirm these results in numerical experiments. Moreover, it remains to explain theoretically the effect of high apparent dimension in the case of extreme events and small catalog size, that was reported in Chapter V.

Analog forecasting uncertainties

In Chapter III, I focused more on average forecasts than forecast covariances, or in a broader sense forecast uncertainty quantification. Yet, the latter is a fundamental aspect of analog forecasting. As stated in Chapter I, analogs are commonly used to produce ensemble forecasts. Even in the pioneering work of Lorenz (1969), analogs were not used to produce forecasts, but to estimate predictability. Atencia and Zawadzki (2017) show that analog ensembles have interesting dynamical properties.

In a formalism similar to the one of Chapter III, I would like to relate three types of uncertainties. The first one is the *forecast error*, which is the difference between the forecast given by analogs (i.e., for instance, an average over the successors) and the true future state, assuming again that the flow is deterministic. The second type of uncertainty is the *empirical uncertainty estimate* that is part of the forecast process (e.g., the covariance of the sucessors). The third type of uncertainty is the *variability of the forecasting process*, which, in the case of a deterministic flow, is a function of the realization of the catalog. Assuming that the catalog is a fixed-length trajectory in phase-space, the catalog is a random variable which depends on the initial point of this trajectory. A particular realization of the catalog is then given by the choice of such an initial point, and this determines the analogs of any given target point. For instance, this variability due to the realization is expressed in the probability distributions of Chapter II.

Limiting behaviour of analog sampling of stochastic processes

In Chapter III, only deterministic dynamics were considered. On the other hand, in Chapter V, analogs were used to emulate the stochastic model equation of the latent-space variable, which was a Markov process. In this case, the objective of analog forecasting is no longer to be as close as possible to *the* true future state, but to *sample* the future state which is inherently a random variable. One can wonder if analog sampling converges to optimal sampling of the initial process in the case of infinitely large catalog. Let us consider the following example.

Assume that X_t is a stochastic Markov process defined by the following discrete-time equation:

$$X_{t+1} = \phi\left(X_t\right) + \eta_t,$$

where ϕ is a deterministic function, and η_t is an i.i.d. additive noise. Then, assume that analogs A_t^k of X_t are taken from a catalog, and that the successors are defined by another

discrete-time equation:

$$\mathrm{A}_{t+1}^k = \tilde{\phi}\left(\mathrm{A}_t^k - \varepsilon_t^k\right) + \tilde{\eta}_t^k + \varepsilon_{t+1}^k \,, \tag{V.8}$$

where $\tilde{\phi}$ is another deterministic function, and $\tilde{\eta}_t$ and ε_t^k are i.i.d. additive noises. $\tilde{\phi}$ represents a difference in dynamics between the analogs and the real process X_t. $\tilde{\eta}_t$ represents a difference in stochasticity between the analogs and the real process X_t. ε_t^k represents an observational noise. Note that, as in Chapter III, I assume noisy but full observations.

Then, assume that the analogs are chosen from the catalog such that

$$\mathrm{A}_t^{k+1} = \mathrm{argmin}_{\mathcal{C}\backslash\left\{\mathrm{A}_t^1, \ldots, \mathrm{A}_t^k\right\}}\left(\|\mathrm{A} - \mathrm{X}_t\|\right) \,,$$

where $\mathcal{C}$ is a catalog generated by iterative applications of Eq. (V.8) from a given initial condition. Analogs are thus minimizers of the distance to X_t. Defining the random variable $r_k = \|\mathrm{A}_t^k - \mathrm{X}_t\|$, the results of Chapter II allow to estimate the probability distribution of this random variable (which have to be slightly modified for the case of extreme events).

Then, the analog sampling $\tilde{\mathrm{X}}_t$ of the real process X_t is given by applying iteratively one of the analog forecasting operators of Chapter III in their multinomial form. Using similar developments as in Chapter III, and under reasonable hypothesis, we should find that:

$$\lim_{(\tilde{\phi},\tilde{\eta},\varepsilon)\to(\phi,\eta,0)} \mathrm{d}\mathbb{P}(\tilde{\mathrm{X}}_t) = \mathrm{d}\mathbb{P}(\mathrm{X}_t) \,,$$

such that the analog sampling converges to the original sampling. It would be interesting to study the rate of convergence, and to look at numerical examples. Also, the case of finite, non-zero observational noise $\varepsilon \neq 0$ should be considered more precisely, to see if analog sampling can still optimally reconstruct the original stochastic process in some sense, and under which hypotheses.

Long-term properties of iterative stochastic analog forecasts

In Chapter III, I focused on what happens to iterative deterministic analog forecasts of a deterministic system. In the previous paragraph, I focused on stochastic analog sampling of a stochastic process. In another research project, I would like to study iterative stochastic analog forecasts of a deterministic system. This is what is done by Yiou (2014)

for atmospheric circulation. I would like to relate the long-term properties of iterative stochastic analog forecasts of deterministic chaotic dynamical systems. The relationship between these long-term analog-based stochastic trajectories and the properties of the original deterministic system would allow to better understand how analog ensemble forecasts of deterministic systems work. The experiments of Yiou (2014) could be repeated in a similar form, but this time on well-known dynamical systems.

Analogs for partial observations

A key aspect of analog forecasting that I did not tackle yet is the case of partial observations. It is essential because it is a very common situation in real applications. Indeed, partial observation data are usually not provided with a physical model for the time evolution of the system state, and therefore analog methods can prove useful. Many works cited in Chapter I make use of Takens embeddings to build an injective mapping from feature space to phase-space, and then use analog forecasting techniques. In this case, can the formalism of Chapter III be used? What would the Jacobian matrix look like? I have no answer to these questions yet. It is likely that using another framework, like the one of Alexander and Giannakis (2020), is the best solution in this case.

Applications

In my thesis, I applied analog forecasting mostly to numerical solutions of idealized dynamical systems. I would like to further apply analog forecasting to real-world cases. In particular, a project on the wind data used in Chapter II is under way. Also, the preliminary results of analog forecasts of ocean waves from point-measurements shown in Chapter III should be continued, with a deeper focus on rogue waves. To my knowledge, the forecast of individual rogue waves from measurements using data assimilation techniques has not been reported in the literature. Using data assimilation to combine different sources of observations (e.g., X-band radar, wave buoys, stereo-video camera) to perform forecasts of rogue waves, possibly combining the methods of Chapter IV and V would be interesting. Finally, applying the methodology of Chapter V to real rain data is a necessary step towards rain forecasts from analogs, since the study of Taillardat et al. (2019) shows that there is room for improvement on the matter.

APPENDIX

Lorenz systems

The three-variable "L63" Lorenz (1963) system of equations is:

$$\begin{cases} \dfrac{\mathrm{d}x_1}{\mathrm{d}t} = \sigma(x_2 - x_1)\,, \\[2em] \dfrac{\mathrm{d}x_2}{\mathrm{d}t} = x_1(\rho - x_3) - x_2\,, \\[2em] \dfrac{\mathrm{d}x_3}{\mathrm{d}t} = x_1 x_2 - \beta x_3\,, \end{cases} \qquad\qquad (\text{V.9})$$

with usual parameters $\sigma = 10$, $\beta = 8/3$ and $\rho = 28$.

The n-variable "L96" Lorenz (1996) system of equations is:

$$\forall i \in [1, n]\,, \quad \frac{\mathrm{d}x_i}{\mathrm{d}t} = -(x_{i-2} + x_{i+1})x_{i-1} - x_i + \theta\,, \qquad\qquad (\text{V.10})$$

where θ is the forcing parameter. We set $n = 8$, $\theta = 8$, and use periodic boundary conditions $x_{i+n} = x_i$.

Product of Hessian with vectors

Let $\mathbf{g}$ a vector-valued, phase-space-dependant function $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$ such as $\mathbf{\Phi}^t$ or $\mathbf{f}$.

The Hessian of $\mathbf{g}$ at $\mathbf{x}$ is noted $\boldsymbol{\nabla}^2\mathbf{g}|_{\mathbf{x}}$. It is of dimension n^3 and its (i, j, k)-th coefficient $\left[\boldsymbol{\nabla}^2\mathbf{g}|_{\mathbf{x}}\right]_{i,j,k}$ equals $\dfrac{\partial^2 g_k}{\partial x_i \partial x_j}(\mathbf{x})$. The product of a Hessian $\boldsymbol{\nabla}^2\mathbf{g}|_{\mathbf{x}}$ with a n-dimensional vector $\mathbf{y}$ is a matrix and its (i, k)-th coefficient $\left[\mathbf{y}\boldsymbol{\nabla}^2\mathbf{g}|_{\mathbf{x}}\right]_{i,k}$ equals $\sum_j y_j \dfrac{\partial^2 g_k}{\partial x_i \partial x_j}$. The double-product of a Hessian with two n-dimensional vectors $\mathbf{y}$ and $\mathbf{z}$ is a vector and its k-th coefficient $\left[\mathbf{y}(\boldsymbol{\nabla}^2\mathbf{g}|_{\mathbf{x}})\mathbf{z}^{\mathrm{T}}\right]_k$ equals $\sum_{i,j} y_j z_i \dfrac{\partial^2 g_k}{\partial x_i \partial x_j}$. The double product of a Hessian $\boldsymbol{\nabla}^2\mathbf{g}|_{\mathbf{x}}$ with two matrices $\mathbf{X}$ and $\mathbf{Y}$ of same shape $K \times n$ is a matrix of shape $k \times n$ and

its (k, j)-th coefficient is $\sum_{l,m} X_{k,l} Y_{k,m} \dfrac{\partial^2 g_j}{\partial x_l \partial x_m}$.

BIBLIOGRAPHY

Thomas AA Adcock and Paul H Taylor. Focusing of unidirectional wave groups on deep water: an approximate nonlinear schrödinger equation-based model. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 465(2110):3083–3102, 2009.

Thomas AA Adcock and Paul H Taylor. The physics of anomalous ('rogue') ocean waves, 2014.

Thomas AA Adcock, Paul H Taylor, S Yan, QW Ma, and PAEM Janssen. Did the draupner wave occur in a crossing sea? *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 467(2134):3004–3021, 2011.

Thomas AA Adcock, Richard H Gibbs, and Paul H Taylor. The nonlinear evolution and approximate scaling of directionally spread wave groups on deep water. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 468(2145): 2704–2721, 2012.

Nail Akhmediev, Adrian Ankiewicz, JM Soto-Crespo, and John M Dudley. Rogue wave early warning through spectral measurements? *Physics Letters A*, 375(3):541–544, 2011.

Mohammad-Reza Alam. Predictability horizon of oceanic rogue waves. *Geophysical Research Letters*, 41(23):8477–8485, 2014.

Romeo Alexander and Dimitrios Giannakis. Operator-theoretic framework for forecasting nonlinear time series with kernel analog techniques. *Physica D: Nonlinear Phenomena*, page 132520, 2020.

Romeo Alexander, Zhizhen Zhao, Eniko Székely, and Dimitrios Giannakis. Kernel analog forecasting of tropical intraseasonal oscillations. *Journal of the Atmospheric Sciences*, 74(4):1321–1342, 2017.

Benetazzo Alvise, Barbariol Francesco, Bergamasco Filippo, Carniel Sandro, and Sclavo Mauro. Space-time extreme wind waves: Analysis and prediction of shape and height. *Ocean Modelling*, 113:201–216, 2017.

Javier Arroyo and Carlos Maté. Forecasting histogram time series with k-nearest neighbours methods. *International Journal of Forecasting*, 25(1):192–207, 2009.

Aitor Atencia and Isztar Zawadzki. A Comparison of Two Techniques for Generating Nowcasting Ensembles . Part II : Analogs Selection and Comparison of Techniques. *Monthly Weather Review*, (2001):2890–2908, 2015. doi: 10.1175/MWR-D-14-00342.1.

Aitor Atencia and Isztar Zawadzki. Analogs on the lorenz attractor and ensemble spread. *Monthly Weather Review*, 145(4):1381–1400, 2017.

A Ayet and P Tandeo. Nowcasting solar irradiance using an analog method and geostationary satellite images. *Solar Energy*, 164:301–315, 2018. ISSN 0038092X. doi: 10.1016/j.solener.2018.02.068.

Jean-Noël Bacro, Carlo Gaetan, Thomas Opitz, and Gwladys Toulemonde. Hierarchical space-time modeling of asymptotically independent exceedances with an application to precipitation data. *Journal of the American Statistical Association*, 115(530):555–569, 2020. doi: 10.1080/01621459.2019.1617152. URL https://doi.org/10.1080/01621459.2019.1617152.

Michael L Banner, Alexander V Babanin, and Ian R Young. Breaking probability for dominant waves on the sea surface. *Journal of Physical Oceanography*, 30(12):3145–3160, 2000.

Cihan Bayındır. Early detection of rogue waves by the wavelet transforms. *Physics Letters A*, 380(1-2):156–161, 2016.

Alvise Benetazzo, Francesco Serafino, Filippo Bergamasco, Giovanni Ludeno, Fabrice Ardhuin, Peter Sutherland, Mauro Sclavo, and Francesco Barbariol. Stereo imaging and x-band radar wave data fusion: An assessment. *Ocean Engineering*, 152:346–352, 2018.

Thomas Bengtsson, Peter Bickel, Bo Li, et al. Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. In *Probability and statistics: Essays*

in honor of David A. Freedman, pages 316–334. Institute of Mathematical Statistics, 2008.

T Brooke Benjamin and James E Feir. The disintegration of wave trains on deep water part 1. theory. *Journal of Fluid Mechanics*, 27(3):417–430, 1967.

T Berry, J R Cressman, and T Sauer. Time-Scale Separation from Diffusion-Mapped Delay Coordinates. *SIAM Journal on Applied Dynamical Systems*, 12(2):618–649, 2013. doi: 10.1137/12088183X.

Tyrus Berry and Timothy Sauer. Local kernels and the geometric structure of data. *Applied and Computational Harmonic Analysis*, 1:1–31, 2015. ISSN 1063-5203. doi: 10.1016/j.acha.2015.03.002. URL http://dx.doi.org/10.1016/j.acha.2015.03.002.

George D Birkhoff. Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences*, 17(12):656–660, 1931.

Simon Birkholz, Carsten Brée, Ivan Veselić, Ayhan Demircan, and Günter Steinmeyer. Ocean rogue waves and their phase space dynamics in the limit of a linear interference model. *Scientific reports*, 6:35207, 2016.

Juliette Blanchet, Simon Stalla, and Jean-Dominique Creutin. Analogy of multiday sequences of atmospheric circulation favoring large rainfall accumulation over the french alps. *Atmospheric Science Letters*, 19(3):e809, 2018.

José C Nieto Borge, Konstanze Reichert, and Katrin Hessner. Detection of spatio-temporal wave grouping properties by using temporal sequences of x-band radar images of the sea surface. *Ocean Modelling*, 61:21–37, 2013.

Rufus Bowen and David Ruelle. The ergodic theory of axiom a flows. In *The theory of chaotic attractors*, pages 55–76. Springer, 1975.

D. S. Broomhead and Gregory P. King. Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena*, 20:217–236, 1986. doi: 10.1016/0167-2789(86)90031-X.

Théophile Caby, Davide Faranda, Giorgio Mantica, Sandro Vaienti, and Pascal Yiou. Generalized dimensions, large deviations and the distribution of rare events. *Physica D: Nonlinear Phenomena*, 400:132143, 2019.

Théophile Caby, Davide Faranda, Sandro Vaienti, and Pascal Yiou. Extreme value distributions of observation recurrences. *arXiv preprint arXiv:2002.10873*, 2020.

Olivier Cappé, Eric Moulines, and Tobias Rydén. *Inference in hidden Markov models.* Springer Science & Business Media, 2006.

Alberto Carrassi, Marc Bocquet, Laurent Bertino, and Geir Evensen. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, 9(5):e535, 2018.

Raymond J Carroll, David Ruppert, Leonard A Stefanski, and Ciprian M Crainiceanu. *Measurement error in nonlinear models: a modern perspective.* CRC press, 2006.

J Cattiaux, R Vautard, C Cassou, P Yiou, Valérie Masson-Delmotte, and F Codron. Winter 2010 in europe: A cold extreme in a warming climate. *Geophysical Research Letters*, 37(20), 2010.

Massimo Cencini, Fabio Cecconi, and Angelo Vulpiani. *Chaos: from simple models to complex systems*, volume 17. World Scientific, 2010.

Dmitry V Chalikov. Linear and nonlinear statistics of extreme waves. *Russian Journal of Numerical Analysis and Mathematical Modelling*, 32(2):91–98, 2017.

Thi Tuyet Trang Chau, Pierre Ailliot, and Valérie Monbet. An algorithm for nonparametric estimation in state-space models. *arXiv preprint 2006.09525*, 2020.

Marios Christou and Kevin Ewans. Examining a comprehensive dataset containing thousands of freak wave events: Part 2—analysis and findings. In *ASME 2011 30th International Conference on Ocean, Offshore and Arctic Engineering*, pages 827–837. American Society of Mechanical Engineers Digital Collection, 2011.

Charles K Chui, Guanrong Chen, et al. *Kalman filtering.* Springer, 2017.

Günther F Clauss and Jan Bergmann. Gaussian wave packets—a new approach to sea-keeping testsof ocean structures. *Applied Ocean Research*, 8(4):190–206, 1986.

William S Cleveland. Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, 74(368):829–836, 1979. doi: 10.2307/2286407.

William S Cleveland and Susan J Devlin. Locally Weighted Regression : An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association*, 83(403):596–610, 1988. doi: 10.2307/2289282.

Stuart Coles, Joanna Bawa, Lesley Trenner, and Pat Dorazio. *An introduction to statistical modeling of extreme values*, volume 208. Springer, 2001.

Richard Courant, Kurt Friedrichs, and Hans Lewy. On the partial difference equations of mathematical physics. *IBM journal of Research and Development*, 11(2):215–234, 1967.

Will Cousins and Themistoklis P Sapsis. Unsteady evolution of localized unidirectional deep-water wave groups. *Physical Review E*, 91(6):063204, 2015a.

Will Cousins and Themistoklis P Sapsis. Reduced order prediction of rare events in unidirectional nonlinear water waves. *arXiv preprint arXiv:1501.05001*, 2015b.

James P Crutchfield and Bruce S McNamara. Equations of motion from a data series. *Complex systems*, 1(417-452):121, 1987.

Anthony C Davison and Raphaël Huser. Statistics of extremes. *Annual Review of Statistics and its Application*, 2:203–235, 2015.

Luca Delle Monache, F Anthony Eckel, Daran L Rife, Badrinath Nagarajan, and Keith Searight. Probabilistic weather prediction with an analog ensemble. *Monthly Weather Review*, 141(10):3498–3516, 2013.

Véronique Ducrocq, François Bouttier, Sylvie Malardel, Thibault Montmerle, and Yann Seity. Le projet arome. *La Houille Blanche*, (2):39–43, 2005.

Guillaume Ducrozet, Félicien Bonnefoy, David Le Touzé, and P Ferrant. 3-d hos simulations of extreme waves in open seas. 2007.

Kristian Dysthe, Harald E Krogstad, and Peter Müller. Oceanic rogue waves. *Annu. Rev. Fluid Mech.*, 40:287–310, 2008.

Davide Faranda, Valerio Lucarini, Giorgio Turchetti, and Sandro Vaienti. Extreme value distribution for singular measures. *arXiv preprint arXiv:1106.2299*, 2011.

Davide Faranda, M Carmen Alvarez-Castro, Gabriele Messori, David Rodrigues, and Pascal Yiou. The hammam effect or how a warm ocean enhances large scale atmospheric predictability. *Nature communications*, 10(1):1–7, 2019.

Mohammad Farazmand and Themistoklis P Sapsis. Reduced-order prediction of rogue waves in two-dimensional deep-water waves. *Journal of Computational Physics*, 340: 418–434, 2017.

J Doyne Farmer and John J Sidorowich. Predicting chaotic time series. *Physical review letters*, 59(8):845, 1987.

Paul Fearnhead and Hans R Künsch. Particle filters and data assimilation. 2018.

Francesco Fedele. Rogue waves in oceanic turbulence. *Physica D: Nonlinear Phenomena*, 237(14-17):2127–2131, 2008.

Francesco Fedele. Geometric phases of water waves. *EPL (Europhysics Letters)*, 107(6): 69001, 2014.

Francesco Fedele, Joseph Brennan, Sonia Ponce De León, John Dudley, and Frédéric Dias. Real world ocean rogue waves explained without the modulational instability. *Scientific reports*, 6:27715, 2016.

Xavier Fettweis, Edward Hanna, Charlotte Lang, Alexandre Belleflamme, Michel Erpicum, and Hubert Gallée. Important role of the mid-tropospheric atmospheric circulation in the recent surface melt increase over the greenland ice sheet. *Cryosphere*, 7: 241–248, 2013.

Jean-François Filipot, Pedro Guimaraes, Fabien Leckler, Jochen Hortsmann, Ruben Carrasco, Étienne Leroy, Nicolas Fady, Mickael Accensi, Marc Prevosto, Rui Duarte, et al. La jument lighthouse: a real-scale laboratory for the study of giant waves and their loading on marine structures. *Philosophical Transactions of the Royal Society A*, 377 (2155):20190008, 2019.

Claude Fischer, Thibaut Montmerle, Loïk Berre, Ludovic Auger, and Simona Ecaterina Ştefănescu. An overview of the variational assimilation in the aladin/france numerical weather-prediction system. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 131 (613):3477–3492, 2005.

George Z Forristall. On the statistical distribution of wave heights in a storm. *Journal of Geophysical Research: Oceans*, 83(C5):2353–2358, 1978.

George Z Forristall. Wave crest distributions: Observations and second-order theory. *Journal of physical oceanography*, 30(8):1931–1943, 2000.

Klaus Fraedrich and Bernd Rückert. Metric adaption for analog forecasting. *Physica A: Statistical Mechanics and its Applications*, 253(1-4):379–393, 1998.

Ana Cristina Moreira Freitas, Jorge Milhazes Freitas, and Sandro Vaienti. Extreme value laws for sequences of intermittent maps. *Proceeding of the American Mathematical Society*, 146(5):2103–2116, 2018. doi: http://dx.doi.org/10.1090/proc/13892.

Jerome H Friedman, Jon Louis Bentley, and Raphael Ari Finkel. An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software (TOMS)*, 3(3):209–226, 1977.

Francesco Fusco and John V Ringwood. Short-term wave forecasting for real-time control of wave energy converters. *IEEE Transactions on sustainable energy*, 1(2):99–106, 2010.

Jürgen Gärtner. On large deviations from the invariant measure. *Theory of Probability & Its Applications*, 22(1):24–39, 1977.

Dimitrios Giannakis. Dynamics-Adapted Cone Kernels. *SIAM Journal on Applied Dynamical Systems*, 14(2):556–608, 2015. doi: 10.1137/140954544.

Dimitrios Giannakis. Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Applied and Computational Harmonic Analysis*, 1:1–59, 2017. ISSN 1063-5203. doi: 10.1016/j.acha.2017.09.001. URL http://dx.doi.org/10.1016/j.acha.2017.09.001.

M Ivette Gomes and Armelle Guillou. Extreme value theory and statistics of univariate extremes: a review. *International statistical review*, 83(2):263–292, 2015.

Zvi Griliches and Vidar Ringstad. Error-in-the-variables bias in nonlinear contexts. *Econometrica: Journal of the Econometric Society*, pages 368–370, 1970.

Ian Grooms. Analog ensemble data assimilation and a method for constructing analogs with variational autoencoders. *arXiv preprint arXiv:2006.01101*, 2020.

Qiuchen Guo and Mohammad-Reza Alam. Prediction of oceanic rogue waves through tracking energy fluxes. In *ASME 2017 36th International Conference on Ocean, Offshore and Arctic Engineering.* American Society of Mechanical Engineers Digital Collection, 2017.

Thomas M Hamill and Jeffrey S Whitaker. Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application. *Monthly Weather Review*, 134(11):3209–3229, 2006.

Franz Hamilton, Tyrus Berry, and Timothy Sauer. Ensemble kalman filtering without a model. *Physical Review X*, 6(1):011021, 2016. doi: 10.1103/PhysRevX.6.011021.

Dieter E Hasselmann, M Dunckel, and JA Ewing. Directional wave spectra observed during jonswap 1973. *Journal of physical oceanography*, 10(8):1264–1280, 1980.

Sverre Haver. A possible freak wave event measured at the draupner jacket january 1 1995. In *Rogue waves*, volume 2004, pages 1–8, 2004.

NICOLAI Haydn and SANDRO Vaienti. Limiting entry times distribution for arbitrary null sets. *arXiv preprint arXiv:1904.08733*, 2019.

Karen L Henderson, D Howell Peregrine, and John W Dold. Unsteady water wave modulations: fully nonlinear solutions and comparison with the nonlinear schrödinger equation. *Wave motion*, 29(4):341–361, 1999.

Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 2020.

Mark P Holland, Renato Vitolo, Pau Rabassa, Alef E Sterk, and Henk W Broer. Extreme value laws in dynamical systems under physical observables. *Physica D: nonlinear phenomena*, 241(5):497–513, 2012.

Paul A Hwang and David W Wang. Directional distributions and mean square slopes in the equilibrium and saturation ranges of the wave spectrum. *Journal of physical oceanography*, 31(5):1346–1360, 2001.

AL Islas and CM Schober. Predicting rogue waves in random oceanic sea states. *Physics of fluids*, 17(3):031701, 2005.

Aglaé Jézéquel, Pascal Yiou, and Sabine Radanovics. Role of circulation in european heatwaves using flow analogues. *Climate dynamics*, 50(3-4):1145–1159, 2018.

Mark Kac. *Probability and related topics in physical sciences*, volume 1. American Mathematical Soc., 1959.

Anatole Katok and Boris Hasselblatt. *Introduction to the modern theory of dynamical systems*, volume 54. Cambridge university press, 1997.

Christian Kharif and Efim Pelinovsky. Physical mechanisms of the rogue wave phenomenon. 2003.

Christian Kharif, Jean-Paul Giovanangeli, Julien Touboul, Laurent Grare, and Efim Pelinovsky. Influence of wind on extreme wave events: experimental and numerical approaches. 2008.

Blair Kinsman. *Wind waves: their generation and propagation on the ocean surface.* Courier Corporation, 1965.

Marco Klein, Matthias Dudek, Günther F Clauss, Sören Ehlers, Jasper Behrendt, Norbert Hoffmann, and Miguel Onorato. On the deterministic prediction of water waves. *Fluids*, 5(1):9, 2020.

John Lamperti. On extreme order statistics. *The Annals of Mathematical Statistics*, 35 (4):1726–1737, 1964.

AL Latifah and E Van Groesen. Coherence and predictability of extreme events in irregular waves. *Nonlinear processes in geophysics*, 19(2), 2012.

G. Lawton. Monsters of the deep. *New Scientist*, 170(2297):28–32, 2001.

Peter D Lax and Robert D Richtmyer. Survey of the stability of linear finite difference equations. *Communications on pure and applied mathematics*, 9(2):267–293, 1956.

Pierre L'Ecuyer, Michel Mandjes, Bruno Tuffin, et al. Importance sampling in rare event simulation., 2009.

Redouane Lguensat, Pierre Tandeo, Pierre Ailliot, Manuel Pulido, and Ronan Fablet. The Analog Data Assimilation. *Monthly Weather Review*, 145(10):4093–4107, 2017. ISSN 0027-0644. doi: 10.1175/MWR-D-16-0441.1.

Michael Selwyn Longuet-Higgins. The statistical analysis of a random, moving surface. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 249(966):321–387, 1957.

Michael Selwyn Longuet-Higgins. Statistical properties of wave groups in a random sea state. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 312(1521):219–250, 1984.

Edward N Lorenz. Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, 20(2):130–141, 1963.

Edward N. Lorenz. Atmospheric Predictability as Revealed by Naturally Occurring Analogues. *Journal of the Atmospheric Sciences*, 26(4):636–646, 1969.

Edward N Lorenz. Predictability: A problem partly solved. In *Proc. Seminar on predictability*, volume 1, 1996.

Valerio Lucarini, Davide Faranda, Jeroen Wouters, and Tobias Kuna. Towards a general theory of extremes for observables of chaotic dynamical systems. *Journal of statistical physics*, 154(3):723–750, 2014.

Valerio Lucarini, Davide Faranda, Jorge Miguel Milhazes de Freitas, Mark Holland, Tobias Kuna, Matthew Nicol, Mike Todd, Sandro Vaienti, et al. *Extremes and recurrence in dynamical systems*. John Wiley & Sons, 2016.

A. K. Magnusson and M. A. Donelan. The andrea wave characteristics of a measured north sea rogue wave. *Journal of Offshore Mechanics and Arctic Engineering*, 135(3): 031108, 2013.

PC Mahalanobis. On the generalised distance in statistics. In *Proceedings of the National Institute of Science of India*, volume 2, pages 49–55, 1936.

AJ Majda, DW McLaughlin, and EG Tabak. A one-dimensional model for dispersive wave turbulence. *Journal of Nonlinear Science*, 7(1):9–44, 1997.

Ivan Markovsky and Sabine Van Huffel. Overview of total least-squares methods. *Signal processing*, 87(10):2283–2302, 2007.

John Milnor. On the concept of attractor. In *The theory of chaotic attractors*, pages 243–264. Springer, 1985.

Hisashi Mitsuyasu, Fukuzo Tasai, Toshiko Suhara, Shinjiro Mizuno, Makoto Ohkusu, Tadao Honda, and Kunio Rikiishi. Observations of the directional spectrum of ocean wavesusing a cloverleaf buoy. *Journal of Physical Oceanography*, 5(4):750–760, 1975.

Grégoire Montavon, Wojciech Samek, and Klaus-Robert Müller. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73:1–15, 2018.

Jérôme Morio, Mathieu Balesdent, Damien Jacquemart, and Christelle Vergé. A survey of rare event simulation methods for static input–output models. *Simulation Modelling Practice and Theory*, 49:287–304, 2014.

Peter Müller, Chris Garrett, and A Osborne. Rogue waves. *Oceanography*, 18(3):66, 2005.

Badrinath Nagarajan, Luca Delle Monache, Joshua P. Hacker, Daran Rife, Keith Searight, Jason C. Knievel, and Thomas N. Nipen. An Evaluation of Analog-Based Postprocessing Methods across Several Variables and Forecast Models. *Weather and Forecasting*, pages 1623–1643, 2015. doi: 10.1175/WAF-D-14-00081.1.

P. Naveau, R. G. Huser, P. Ribereau, and A. Hannart. Modelling jointly low, moderate and heavy rainfall intensities without a threshold selection. *Water Resources Research*, 52:2753–2769, 2016.

Philippe Naveau, Armelle Guillou, and Théo Rietsch. A non-parametric entropy-based approach to detect changes in climate extremes. *Journal of the Royal Statistical Society B.*, 2014.

Bárður A Niclasen, Knud Simonsen, and Anne Karin Magnusson. Wave forecasts and small-vessel safety: A review of operational warning parameters. *Marine structures*, 23 (1):1–21, 2010.

C Nicolis. Atmospheric Analogs and Recurrence Time Statistics: Toward a Dynamical Formulation. *Journal of the Atmospheric Sciences*, 55:465–475, 1998.

C Nicolis, P Perdigao, and S Vannitsem. Dynamics of Prediction Errors under the Combined Effect of Initial Condition and Model Errors. *Journal of the Atmospheric Sciences*, 66:766–778, 2009. doi: 10.1175/2008JAS2781.1.

I Nikolkina, I Didenkulova, E Pelinovsky, and P Liu. Rogue waves in 2006-2010. *Natural Hazards & Earth System Sciences*, 11(11), 2011.

Frédéric Nouguier, Stéphan T Grilli, and Charles-Antoine Guérin. Nonlinear ocean wave reconstruction algorithms based on simulated spatiotemporal data acquired by a flash lidar camera. *IEEE Transactions on Geoscience and Remote Sensing*, 52(3):1761–1771, 2013.

FWJ Olver. Error bounds for the laplace approximation for definite integrals. *Journal of Approximation Theory*, 1(3):293–313, 1968.

N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw. Geometry from a Time Series. *Physical Review Letters*, 45(9):712–716, 1980. doi: 10.1103/PhysRevLett.45.712.

E Pelinovsky, E Shurgalina, and N Chaikovskaya. The scenario of a single freak wave appearance in deep water–dispersive focusing mechanism framework. *Natural Hazards and Earth System Sciences*, 11(1):127–134, 2011.

D Howell Peregrine. Interaction of water waves and currents. In *Advances in applied mechanics*, volume 16, pages 9–117. Elsevier, 1976.

OM Phillips, Daifang Gu, and Mark Donelan. Expected structure of extreme waves in a gaussian sea. part i: Theory and swade buoy measurements. *Journal of Physical Oceanography*, 23(5):992–1000, 1993.

Nick Pizzo and W Kendall Melville. Focusing deep-water surface gravity wave packets: wave breaking criterion in a simplified model. *Journal of Fluid Mechanics*, 873:238–259, 2019.

P Platzer, P Yiou, P Tandeo, P Naveau, and J-F Filipot. Using local dynamics to explain analog forecasting of chaotic systems. *Journal of the Atmospheric Sciences*, 2020, under revision.

Paul Platzer, Pascal Yiou, Pierre Tandeo, Philippe Naveau, and Jean-François Filipot. Predicting analog forecasting errors using dynamical systems. In *CI 2019: 9th International Workshop on ClimateInformatics*, 2019.

Henri Poincaré. Sur le problème des trois corps et les équations de la dynamique. *Acta mathematica*, 13(1):A3–A270, 1890.

Yves Quilfen, Maria Yurovskaya, Bertrand Chapron, and Fabrice Ardhuin. Storm waves focusing and steepening in the agulhas current: Satellite observations and modeling. *Remote sensing of Environment*, 216:561–571, 2018.

Sidney I Resnick. *Heavy-tail phenomena: probabilistic and statistical modeling.* Springer Science & Business Media, 2007.

Yoann Robin, Pascal Yiou, and Philippe Naveau. Detecting changes in forced climate attractors with Wasserstein distance. *Nonlinear Processes in Geophysics*, 24(3):393–405, jul 2017. ISSN 1607-7946. doi: 10.5194/npg-24-393-2017. URL https://www.nonlin-processes-geophys.net/24/393/2017/.

Viktor Petrovich Ruban. Predictability of the appearance of anomalous waves at sufficiently small benjamin–feir indices. *JETP letters*, 103(9):568–572, 2016.

Andrea Ruju, Jean-François Filipot, Abderrahim Bentamy, and Fabien Leckler. Spectral wave modelling of the extreme 2013/2014 winter storms in the north-east atlantic. *Ocean Engineering*, accepted.

Suranjana Saha, Shrinivas Moorthi, Hua-Lu Pan, Xingren Wu, Jiande Wang, Sudhir Nadiga, Patrick Tripp, Robert Kistler, John Woollen, David Behringer, et al. The ncep climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, 91 (8):1015–1058, 2010.

EV Sanina, SA Suslov, D Chalikov, and AV Babanin. Detection and analysis of coherent groups in three-dimensional fully-nonlinear potential wave fields. *Ocean Modelling*, 103: 73–86, 2016.

Tim Sauer. Time series prediction by using delay coordinate embedding. *Time Series Prediction: forecasting the future and understanding the past*, 1994.

Tim Sauer, James A Yorke, and Martin Casdagli. Embedology. *Journal of Statistical Physics*, 65(3):579–616, 1991. doi: 10.1007/BF01053745. URL https://doi.org/10.1007/BF01053745.

F Schenk and E Zorita. Reconstruction of high resolution atmospheric fields for northern europe using analog-upscaling. *Climate of the Past*, 8(5):1681, 2012.

CJE Schuurmans and SCHUURMANS CJE. A 4-year experiment in long-range weather forecasting, using circulation analogues. 1973.

Florian Sévellec and Sybren S Drijfhout. A novel probabilistic forecast system predicting anomalously warm 2018-2022 reinforcing the long-term global warming trend. *Nature communications*, 9(1):1–12, 2018.

J. N. Sharma. Development and evaluation of a procedure for simulating a random directional second-order sea surface and associated wave forces. 1980.

Y Sinai. *Introduction to ergodic theory*. Princeton University Press, 1977.

Yakov G Sinai. Gibbs measures in ergodic theory. *Russian Mathematical Surveys*, 27(4): 21, 1972.

Alexey Slunyaev. Predicting rogue waves. *Moscow University Physics Bulletin*, 72(3): 236–249, 2017.

Alexey Slunyaev, Efim Pelinovsky, and C Guedes Soares. Modeling freak waves from the north sea. *Applied Ocean Research*, 27(1):12–22, 2005.

George Sugihara. Nonlinear forecasting for the classification of natural time series. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*, 348(1688):477–495, 1994.

M. Taillardat, A-L. Fougères, P. Naveau, and O. Mestre. Forest-based and semiparametric methods for the postprocessing of rainfall ensemble forecasting. *Weather and Forecasting*, pages 617–633, 2019. doi: https://doi.org/10.1175/WAF-D-18-0149.1.

Floris Takens. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick 1980*, pages 366–381. Springer, 1981. doi: https://doi.org/10.1007/BFb0091924.

Pierre Tandeo, Pierre Ailliot, Juan Ruiz, Alexis Hannart, Bertrand Chapron, Anne Cuzol, Valérie Monbet, Robert Easton, and Ronan Fablet. Combining analog method and ensemble data assimilation: application to the lorenz-63 chaotic system. In *Machine learning and data mining approaches to climate science*, pages 3–12. Springer, 2015.

Pierre Tandeo, Aitor Atencia, and Cristina Gonzalez Haro. Combining 15 years of microwave sst and along-track ssh to estimate ocean surface currents. In *CI 2016: 6th International Workshop on Climate Informatics*, pages 1–4, 2016.

Pierre Tandeo, Pierre Ailliot, Marc Bocquet, Alberto Carrassi, Takemasa Miyoshi, Manuel Pulido, and Yicun Zhen. Joint estimation of model and observation error covariance matrices in data assimilation: a review. *arXiv preprint arXiv:1807.11221*, 2018.

Matthew Thorpe, Serim Park, Soheil Kolouri, Gustavo K Rohde, and Dejan Slepčev. A transportation

$$l^p$$

distance for signal analysis. *Journal of mathematical imaging and vision*, 59(2):187–210, 2017.

Michael K. Tipett and Timothy DelSole. Constructed Analogs and Linear Regression. *Monthly Weather Review*, 141:2519–2525, 2013. doi: 10.1175/MWR-D-12-00223.1.

Vasily V Titov, Frank I Gonzalez, EN Bernard, Marie C Eble, Harold O Mofjeld, Jean C Newman, and Angie J Venturato. Real-time tsunami forecasting: Challenges and solutions. *Natural Hazards*, 35(1):35–41, 2005.

Hendrik L Tolman et al. User manual and system documentation of wavewatch iii tm version 3.14. *Technical note, MMAB Contribution*, 276:220, 2009.

Julien Touboul and Christian Kharif. Effect of vorticity on the generation of rogue waves due to dispersive focusing. *Natural Hazards*, 84(2):585–598, 2016.

Gwladys Toulemonde, Armelle Guillou, and Philippe Naveau. Particle filtering for gumbel-distributed daily maxima of methane and nitrous oxide. *Environmetrics*, 24 (1):51–62, 2013.

H M Van Den Dool. Searching for analogues , how long must we wait ? *Tellus A: Dynamic Meteorology and Oceanography*, 46(3):314–324, 1994. doi: 10.3402/tellusa.v46i3.15481.

Huug Van den Dool, Principal Scientist Cpc, and Huug Van Den Dool. *Empirical methods in short-term climate prediction*. Oxford University Press, 2007.

Peter Jan Van Leeuwen. Particle filtering in geophysical systems. *Monthly Weather Review*, 137(12):4089–4114, 2009.

Divakar Viswanath. Lyapunov exponents from random fibonacci sequences to the loren-zequations. Technical report, Cornell University, 1998.

Xinyang Wang, Joanna Slawinska, and Dimitrios Giannakis. Extended-range statistical enso prediction through operator-theoretic techniques for nonlinear dynamics. *Scientific reports*, 10(1):1–15, 2020.

Fredrik Wetterhall, Sven Halldin, and Chong-yu Xu. Statistical precipitation downscaling in central sweden with the analogue method. *Journal of Hydrology*, 306(1-4):174–190, 2005.

Robert L Wolpert. Stationary gamma processes. 2016.

Prince K Xavier and BN Goswami. An analog method for real-time forecasting of summer monsoon subseasonal variability. *Monthly Weather Review*, 135(12):4149–4160, 2007.

Pascal Yiou. AnaWEGE: a weather generator based on analogues of atmospheric circulation. *Geoscientific Model Development*, 7(2):531–543, apr 2014. ISSN 1991-9603. doi: 10.5194/gmd-7-531-2014.

Pascal Yiou and Céline Déandréis. Stochastic Ensemble Climate Forecast with an Analogue Model. *Geoscientific Model Development*, (September):1–18, 2018.

Pascal Yiou, Tamara Salameh, Philippe Drobinski, Laurent Menut, Robert Vautard, and Mathieu Vrac. Ensemble reconstruction of the atmospheric column from surface pressure using analogues. *Climate dynamics*, 41(5-6):1333–1344, 2013.

Lai-Sang Young. Dimension, entropy and lyapunov exponents. *Ergodic theory and dynamical systems*, 2(1):109–124, 1982.

Henry C Yuen and Bruce M Lake. Nonlinear dynamics of deep-water gravity waves. In *Advances in applied mechanics*, volume 22, pages 67–229. Elsevier, 1982.

VE Zakharov, SV Manakov, SP Novikov, and LP Pitaevskii. Theory of solitons: inverse problem method, 1980.

Vladimir E Zakharov. Stability of periodic waves of finite amplitude on the surface of a deep fluid. *Journal of Applied Mechanics and Technical Physics*, 9(2):190–194, 1968.

Zhizhen Zhao and Dimitrios Giannakis. Analog forecasting with dynamics-adapted kernels. *Nonlinearity*, 29(9):2888–2939, 2016. doi: 10.1088/0951-7715/29/9/2888.

Yicun Zhen, Pierre Tandeo, Stéphanie Leroux, Sammy Metref, Thierry Penduff, and Julien Le Sommer. An adaptive optimal interpolation based on analog forecasting: application to ssh in the gulf of mexico. *Journal of Atmospheric and Oceanic Technology*, 37(9):1697–1711, 2020.

Jakob Zscheischler, Philippe Naveau, Olivia Martius, Sebastian Engelke, and Christoph C Raible. Evaluating the dependence structure of compound precipitation and wind speed extremes. *Earth System Dynamics Discussions*, pages 1–23, 2020.

POSTFACE

Je voudrais qu'à ma mort, avant de dire "c'était un grand artiste", les gens disent "il nous a apporté telle chose, il a fait telle chose de bien". Ça me paraît beaucoup plus important que de dire "oh, qu'est-ce qu'il chantait bien!".
  – Kery James, *Interview Street Press (Octobre 2019)*

Au commencement de ma thèse, j'avais l'esprit plein d'idées reçues, de certitudes plates mais rassurantes. Je voyais la recherche scientifique comme une entreprise désintéressée d'approfondissement de la connaissance, dont le but était d'apporter des outils de compréhension et de maîtrise à l'humanité. Je ne vais pas retracer ici l'ensemble des évènements personnels qui m'ont conduit à changer d'opinion. Je ne vais pas, non plus, faire un exposé détaillé en faveur des positions que je défends aujourd'hui. Mais je tiens à donner quelques pistes pour des personnes qui se trouveraient face aux mêmes types de questionnements que ceux j'ai rencontré. Puisse une âme perdue (en plein doctorat?) trouver ici des débuts de réponses à ses questions.

Les sciences des données vont-elles sauver la planète ? Jérémy et moi avons tenté de donner notre opinion à ce sujet dans une tribune[1]. En allant plus loin, on peut même se demander si les sciences du climat sauveront le climat ? Si la parole des scientifiques du climat est importante, il est nécessaire de se détacher d'une vision simpliste dans laquelle les scientifiques du climat seraient des lanceurs d'alertes, sauveurs de l'humanité. Je recommande la lecture de "L'Évènement Anthropocène" de Christophe Bonneuil et Jean-Baptiste Fressoz, qui permet de prendre du recul par rapport au discours des "anthropocénologues". Comprendre l'histoire des alertes environnementales, de leur déni, et des mécanismes sociaux, techniques, économiques et idéologiques qui nous ont amené où nous sommes aujourd'hui, est une étape essentielle pour faire face aux ravages écologiques.

De manière plus générale, on peut s'interroger sur les sciences de la nature, et leur prétendue neutralité. Un texte de l'hebdomadaire lundi matin[2] y réfléchit. À la lecture

---

1. "Le big data ne sauvera pas la planète", Libération, publié le 15 octobre 2019.
2. https://lundi.am/Scientifiques-de-la-nature-vous-etes-des-sujets-politiques

d'une histoire des sciences et des savoirs[3], l'idée d'une science neutre paraît naïve, tant l'imbrication historique entre science, gouvernance, et guerre, y apparaît évidente. Dans *Les enjeux politiques de la science*[4], John Stewart montre que même la méthode scientifique n'est pas neutre, en tant qu'elle répond d'abord à un recherche de puissance et de maîtrise. La science s'accompagne également d'une idéologie, le scientisme, vivement critiquée[5].

Peut-on changer le monde de la recherche de l'intérieur ? S'il est toujours possible de se battre pour de meilleures conditions de travail et moins de concurrence, de porter les voix des travailleuses et travailleurs[6], il semblerait qu'un changement en profondeur soit quasi impossible. C'est pourquoi certains appellent à déserter la recherche scientifique[7], comme d'autres désertent le milieu bancaire[8]. La question de continuer, ou non, la recherche scientifique, n'est pas nouvelle[9] et continue de traverser le champ universitaire[10].

Ma thèse est en lien, bien qu'assez indirect, avec le développement d'infrastructures de captation d'énergies renouvelables. On peut également se demander si ces infrastructures nous sauveront du problème climatique ? L'idée d'une transition énergétique, et même l'idée d'une transition tout court[11], pourraît bien être un piège dans lequel les écologistes feraient mieux de ne pas tomber. De même, les énergies renouvelables ne doivent pas à mon avis être vues comme une "solution", mais comme de nouvelles technologies reposant sur des infrastructures techniques, énergétiques, et sociales, qu'il s'agit d'analyser pour comprendre leurs potentiels et leurs limites.

La lecture de ces quelques paragraphes pourraît donner une impression pessimiste, voire défaitiste à l'égard de la science. Au contraire, toutes ces critiques de l'intérieur, toutes ces réflexions sur les sciences et les techniques sont porteuses d'un autre regard

---

3. *Histoire des sciences et des savoirs*, sous la direction de Dominique Pestre. Bien qu'un peu hardus pour des non-spécialistes, les trois tomes sont intéressants pour comprendre d'où l'on vient en tant que scientifique.

4. Impascience, n°1, disponible sur `http://science-societe.fr/`.

5. *La Nouvelle Église Universelle*, A. Grothendieck, Survivre & Vivre n°9, dispo sur `http://science-societe.fr/`

6. Voir *Lucioles* `http://indiscipline.fr/lucioles-numero-2/`.

7. `https://lundi.am/Avis-aux-chercheurs-aux-professeurs-aux-ingenieurs`

8. *Faire sauter la banque*, Jérémy Désir, éditions Divergences

9. `https://www.cairn.info/revue-ecologie-et-politique-2016-1-page-159.htm`

10. https://ecopolien.hypotheses.org/seminaires-debats/seminaire-debat-allons-nous-continuer-la-recherche-scientifique

11. Voir le livre *Écologie sans transition* du collectif Désobéissance Écolo Paris, éditions Divergences.

sur les savoirs, d'un autre rapport au monde qui constitue, à mon sens, le germe d'une transformation nécessaire. Les initiatives telles que les ateliers d'écologie politique[12] fleurissent partout. L'histoire de l'auto-critique de la science est riche (Survivre et Vivre, Labo Contestation, Impascience). Le monde universitaire est aussi un espace de lutte, d'organisation du rapport de force social, comme en témoignent les récentes mobilisations contre la LPPR, et de nombreuses autre initiatives[13]. Les scientifiques peuvent tout à fait aider l'humanité, tant qu'ils se posent les bonnes questions...

---

12. À Toulouse `https://atecopol.hypotheses.org`, à Paris `https://ecopolien.hypotheses.org`, à Grenoble `https://campus1point5grenoble.org/?page_id=132`, à Rennes `https://epolar.hypotheses.org/`, et bientôt à Brest ?

13. http://science-societe.fr/appel-a-une-occupation-numerique-dans-lenseignement-superieur-et-la-recherche-les-1er-et-2-fevrier-2020, par exemple.

Titre : Prévision par analogues de systèmes dynamiques : application à des variables géophysiques et aux vagues océaniques

Mot clés : prévision, analogues, systèmes dynamiques, geophysique, vagues, extrêmes

Résumé : La variété des méthodes de prévision des variables géophysiques va croissante. Il est donc essentiel de comprendre les propriétés, les avantages et les limites de chaque méthode. Nous étudions les "analogues", utilisés en météorologie depuis plus de 60 ans. La probabilité de trouver de bons analogues est exprimée via des distributions de probabilité de distances d'analogues, dont la forme analytique est obtenue pour les systèmes dynamiques ergodiques. L'application aux données de vent à dix mètres montre que les premières distances d'analogues sont très similaires, ce qui n'est pas le cas pour les systèmes de faible dimension. Ensuite, nous comparons l'efficacité de plusieurs méthodes de prévision par analogues, en reliant les erreurs de prévision au flot du système. L'influence du bruit d'observation sur les prévisions par analogues est étudiée théoriquement et numériquement, justifiant l'utilisation d'un grand nombre d'analogues pour atténuer l'effet du bruit. L'applicabilité des prévisions par analogues aux variables aléatoires à queue lourde est testée numériquement sur un modèle espace-état, montrant la nécessité de disposer d'un plus grand nombre de données pour prévoir les événements extrêmes. Enfin, un modèle physique de prévision des vagues océaniques extrêmes est proposé. Celui-ci est basé uniquement sur des mesures de vitesse de crête, afin de simplifier la prévision.

Title: Forecasts of dynamical systems from analogs: applications to geophysical variables with a focus on ocean waves

Keywords: forecast, analogs, dynamical systems, geophysical, ocean waves, extremes

Abstract: The variety of forecasting methods for geophysical variables is increasing. Therefore, understanding the properties, advantages and limitations of each method is crucial. We focus on "analogs" which have been used in meteorology for more than 60 years. The probability to find good analogs is expressed through closed-form probability distributions of analog-to-target distances, derived for ergodic dynamical systems. Application to 10 m-wind data shows that the first analog-to-target distances are very similar, which is not the case for low-dimensional systems. Then, we compare the efficiency of several analog forecasting methods, by linking analog forecasting errors to the flow map of the system. The influence of observational noise on analog forecasts is studied theoretically and numerically, justifying the use of a large number of analogs to mitigate the effect of noise. The applicability of analog forecasts to heavy-tailed random variables is tested numerically on a state-space model, witnessing the need for a larger amount of data to forecast extreme events. Finally, a physics-based methodology is proposed to forecast extreme ocean waves, using only crest velocity measurements in order to simplify the forecasting process.