Rapport de projet

Projet de Statistiques Appliquées

Montée des océans

Réalisé par :
Audrey Bovet
Léo Dony
Pierrette Josiane Makamwe
Alexandre Nonnenmacher

17 mai 2025

Table des matières

Remerciements

Nous tenons à exprimer notre gratitude à toutes les personnes qui nous ont accompagnés tout au long de ce projet de statistiques appliquées.

Nous remercions tout particulièrement M. Bilal Al Taki pour son encadrement attentif et bienveillant ainsi que ses conseils tout au long du projet.

Nous remercions également M. Olivier Guéant, notre référent pédagogique, pour les retours apportés lors du rendu intermédiaire et pour le temps qu'il consacre à l'évaluation de notre travail.

Nous adressons également nos remerciements à M. Étienne Donier-Méroz, M. Paul Guillot et M. Félix Pasquier, coordinateurs du projet, pour leur organisation et pour la mise en place de ce projet qui nous a permis d'approfondir nos compétences en statistiques appliquées..

Introduction

0.1 Contexte et problématique

Depuis le début de l'ère industrielle, les activités humaines - en particulier via l'accumulation progressive de gaz à effet de serre qui en résultent - ont provoqué un réchauffement global et inédit du climat terrestre, dont les conséquences sont désormais observables sur l'ensemble des systèmes naturels et humains. La montée du niveau des eaux est aujourd'hui l'un des phénomènes les plus préoccupants liés au réchauffement climatique. À l'instar d'autre phénomènes climatiques, la montée des eaux s'accélère à un rythme important et s'impose alors comme un phénomène concret et mesurable, après avoir longtemps été reléguée à un horizon lointain. Ses conséquences sont majeures, tant pour les sociétés humaines que pour les écosystèmes marins et les littoraux, profondément impactés et bouleversés par ces transformations. Comprendre et anticiper cette évolution n'est plus une simple question scientifique, mais bien une nécessité pour l'adaptation des territoires côtiers, la gestion des risques et la préservation des populations et écosystèmes les plus vulnérables.

Sur le plan physique, l'élévation du niveau marin est principalement due au réchauffement climatique qui agit de plusieurs façons complémentaires, les deux effets principaux étant d'une part, la dilatation thermique des océans – les masses d'eau se dilatant sous l'effet de la chaleur – et d'autre part la fonte accélérée des calottes glaciaires et des glaciers continentaux.

Les enjeux associés à cette montée des eaux sont nombreux. Sur le plan humain, des millions de personnes vivant dans les zones côtières les plus sensibles pourraient être déplacées d'ici la fin du siècle, et certaines nations telles que les îles Kiribati pourraient tout simplement disparaître. Des mégapoles comme Jakarta, New York ou Lagos sont directement menacées. Les risques d'érosion du littoral, d'inondations récurrentes et de salinisation des nappes phréatiques viennent compromettre la sécurité alimentaire, l'accès à l'eau potable et l'intégrité des infrastructures. Sur le plan environnemental, les écosystèmes marins et côtiers subissent également de fortes pressions : la submersion des zones humides, la modification des habitats littoraux ou bien l'acidification des océans fragilisent les chaînes alimentaires et la biodiversité. Ces perturbations peuvent entraîner la disparition d'espèces (30% des récifs coralliens ont disparu depuis 1980, par exemple) et perturber les équilibres biologiques et les services écosystémiques essentiels.

0.2 Objectifs du projet

L'objectif de ce projet de statistique et science des données appliquées est d'identifier les causes et les facteurs responsables de la hausse du niveau des eaux, afin de pouvoir dans un second temps modéliser cette évolution, dans le but de prédire les effets futurs du changement climatique sur le niveau marin. Nous avons décidé de concentrer notre travail sur la zone géo-

graphique de la Méditerranée. Cette région nous paraît intéressante à étudier car on y observe la plupart des effets décrits classiquement dans la littérature sur la montée des eaux, mais aussi certains effets plus spécifiques à la région. C'est également une région très touristique, ce qui donne une importance économique et politique aux résultats que nous pourrons obtenir. Notre travail s'articulera donc autour de la question de recherche suivante : « Quels sont les facteurs principaux contribuant à la montée du niveau des eaux en Méditerranée, et comment peut-on prédire son évolution future ? ».

Notre travail se découpe en quatre grands axes. Nous avons commencé par effectuer une revue de littérature afin de faire un état de l'art pour identifier les facteurs influençant la montée du niveau des eaux mais aussi pour comprendre les modèles et méthodes statistiques et d'apprentissage automatique (machine learning) utilisées par les études passées sur le sujet. Nous sommes ensuite passés par une phase d'analyse exploratoire sur nos données afin d'avoir une compréhension plus fine des phénomènes physiques à l'œuvre. Suite à cela, nous avons pu nous atteler à la modélisation et à la prédiction en implémentant les méthodes trouvées précédemment et en les appliquant à nos données. La dernière partie consiste en l'interprétation de nos résultats, en nous comparant aux résultats d'autres projets issus d'organismes scientifiques.

0.3 Méthodologie adoptée

Nous allons présenter dans cette section les différentes étapes du projet ainsi que les outils et méthodes que nous allons utiliser pour répondre à notre problématique.

Pour la revue de littérature, nous avons étudié des articles scientifiques afin de nous familiariser avec les mécanismes de la montée des eaux et d'identifier les facteurs. Nous avons complété ces recherches par l'étude de projets déjà menés sur le sujet, ce qui nous a également permis de découvrir les moyens techniques mis en œuvre dans ce genre de projets. Nous avons également passé en revue des sites de partage de projets tels que GitHub ou Kaggle pour améliorer notre connaissance des modèles et méthodes utilisées.

Lors de l'analyse exploratoire de nos données, nous avons rencontré une difficulté majeure qui a freiné toute la suite du projet. En effet, nous n'avions pas accès à une base de données ad hoc et nous avons donc dû la créer de toutes pièces par nous même. Cette création nous a fait perdre beaucoup de temps car nous avons eu à parcourir de nombreux sites proposant des données climatiques, tels que ceux de la NASA, l'ESA ou encore Copernicus afin d'identifier des jeux de données pertinents et compatibles, que nous avons dû ensuite fusionner entre eux pour obtenir notre base de données. Une fois cette base constituée, nous avons utilisé Python afin de visualiser nos données, d'effectuer certaines corrections et de vérifier empiriquement les liens entre nos variables explicatives et notre variable d'intérêt.

Nous avons ensuite continué le projet sur Python pour la modélisation et la prédiction. Nous avons implémenté différentes méthodes de régression en prenant en compte les spécificités de nos données au fur et à mesure.

Chapitre 1

Revue de littérature

D'après les différents rapports du GIEC [11] et d'autres études scientifiques ([13], [3], [6], [12]), le niveau des océans est resté relativement stable au cours des deux ou trois derniers millénaires, avec une élévation moyenne inférieure à 0,5 mm/an. Cependant, le XXe siècle a marqué le début d'une accélération du rythme de la montée des eaux, avec une élévation moyenne autour de 1,6 à 1,8 mm/an durant le siècle. Depuis 1993, les observations satellites montrent une augmentation bien plus rapide de la hausse, avec une moyenne au-delà de 3,5 mm/an au cours des trois dernières décennies. Les projections du GIEC indiquent que cette tendance à la hausse va se poursuivre, les différents scénarios estimant une élévation annuelle de 5 à 10 mm d'ici 2050.

1.1 Les facteurs de la montée des eaux

Les deux facteurs les plus impactants sur le niveau des océans sont la dilatation thermique et la fonte des glaces [9], [7]. À eux deux, ils sont responsables de plus de 90% de l'augmentation du niveau des eaux.

1.1.1 La dilatation thermique

L'eau a la propriété de se dilater lorsqu'elle se réchauffe. Puisque les océans stockent environ 90% de l'excédent de chaleur dû au réchauffement climatique, cet effet joue un rôle clé dans l'élévation du niveau des océans. Selon différents travaux de recherche, la dilatation thermique a contribué à 30-40% de l'élévation totale du niveau des océans depuis le début du siècle dernier. Sa contribution actuelle est estimée entre 1,08 et 1,72 mm/an [9].
Puisque le coefficient de dilatation thermique de l'eau dépend à la fois de la température et de la pression, l'effet n'est pas homogène sur le globe, et il varie selon la profondeur. Le coefficient est en général plus fort à la surface puis diminue jusqu'à 1 km de profondeur avant d'augmenter progressivement jusqu'au fond des océans. Il prend des valeurs typiquement comprises entre 1 et 2,5 ppm/K (un coefficient de 1 ppm/K signifie que pour un Kelvin de température supplémentaire, le volume augmente d'un millionième).
Du fait d'un coefficient élevé en surface, les couches superficielles de l'océan répondent de manière rapide aux variations climatiques, augmentant à court terme le niveau des océans. À long terme, l'élévation du niveau des océans va se poursuivre pendant des siècles, même en cas d'une stabilisation des températures atmosphériques, car les variations du coefficient de dilatation sont telles que les couches plus profondes de l'océan ont un temps de réponse bien plus long et donc la chaleur va continuer à pénétrer jusqu'au fond de l'océan par inertie thermique.

Cet effet, dit stérique [14], [20], correspond à une augmentation de volume à masse constante, contrairement aux autres effets, dont les contributions sont barystatiques, c'est-à-dire qu'elles correspondent à une variation de la masse d'eau dans les océans.

1.1.2 La fonte des glaces

Le réchauffement climatique entraîne une fonte accélérée des glaciers, calottes polaires et autres masses de glaces, ajoutant directement de l'eau dans les océans. Du fait de la quantité très importante d'eau contenue dans ces différentes masses glacées, ce phénomène joue un rôle majeur dans l'élévation du niveau des océans. Les différents travaux sur le sujet montrent que la fonte des glaces a contribué à environ 60% de l'élévation totale du niveau des océans depuis les années 1900. Sa contribution actuelle est estimée aux alentours de 2 mm/an. [7], [3]
Lorsqu'on parle de la fonte des glaces, on considère essentiellement la fonte des glaces continentales. La fonte des glaces flottantes joue également un rôle dans la montée du niveau des océans, mais très largement négligeable devant celui des glaces continentales (environ 10^5 fois moins important selon une étude menée par des chercheurs de l'ENS Lyon, [7]).
Il faut également différencier les effets des glaciers de montagnes et calottes glaciaires de ceux des inlandsis (couvertures de glace de l'échelle d'un continent et très épaisses, cf. Groenland et Antarctique), dont les contributions ne sont pas les mêmes à court et à long terme.
Les glaciers de montagnes et les calottes glaciaires représentent moins de 0,5% du volume des glaces terrestres, mais leur contribution à court terme est très importante puisqu'ils fondent bien plus rapidement que les inlandsis, du fait de leur taille et de leur position dans des zones moins froides. D'après une synthèse du GIEC de 2019 [10], leur contribution est située entre 0,53 et 0,69 mm/an au XXIe siècle. Depuis le début du XXe siècle, la fonte des calottes glaciaires et des glaciers de montagne serait responsable d'environ 1/3 de l'élévation observée du niveau de la mer (soit plus de la moitié de la contribution imputée à la fonte des glaces). A long terme, la contribution des glaciers et calottes glaciaires sera bien plus faible du fait qu'ils auront presque entièrement fondu d'ici 2 siècles dans les scénarios les plus pessimistes. Si l'ensemble des glaciers et calottes fondait, leur contribution totale serait estimée entre 0,15 et 0,5 m, alors que la fonte totale des inlandsis du Groenland et de l'Antarctique amènerait à une élévation d'environ 60 à 70 m.
La fonte des inlandsis est la plus préoccupante à long terme. Malgré une part en augmentation ces dernières années, la fonte des inlandsis du Groenland et de l'Antarctique n'a encore qu'une contribution modérée à la montée du niveau des eaux. Mais à long terme la fonte de ces 3 inlandsis (1 au Groenland, 2 en Antarctique) sera la cause majeure de la montée du niveau des océans, puisqu'ils contiennent environ 30 millions de km^3 de glace, ce qui représente 60 à 70 m d'élévation s'ils fondaient totalement (Cela prendrait plusieurs milliers d'années même dans le scénario le plus pessimiste).
Dans le détail, l'inlandsis du Groenland est celui dont la contribution s'est le plus accélérée, passant de 15% à 25% de la contribution totale entre 1993 et aujourd'hui, l'inlandsis Antarctique Ouest est celui pour lequel l'instabilité à court terme est la plus forte, avec des glaciers qui fondent de plus en plus vite, et l'inlandsis Antarctique Est, le plus grand des trois (80% du volume total), est relativement stable à court terme mais représente un danger majeur à long terme.
A cela s'ajoutent des phénomènes de rétroaction, qui viennent accélérer la fonte des glaces. Les principaux facteurs de rétroaction sont les suivants :

- Rétroaction fonte-albédo : La neige a un albédo très élevé, et quand elle fond, elle laisse sa place à de l'eau liquide ou à de la glace (si regel), qui ont un albédo plus faible, ce qui a pour effet d'accélérer la fonte car la température augmente plus rapidement au soleil.
- Rétroaction altitude-fonte : Dans le cas des inlandsis, lors de la fonte, la surface de glace

se retrouve à une altitude plus basse, et est donc exposée à des températures de l'air plus élevées.

1.1.3 Les facteurs secondaires

A ces 2 facteurs principaux s'ajoutent des facteurs secondaires, dont les effets sont plus faibles et potentiellement différenciés selon les régions où l'on se place. Voici quelques-uns des facteurs secondaires les plus souvent mentionnés dans la littérature.

L'exploitation et la surexploitation des ressources aquifères (pour l'agriculture et la consommation par exemple), entraîne un rejet d'eau douce dans les océans. Cet effet a contribué à environ 10% de la hausse observée selon une étude de Léonard Konikow menée en 2011.

L'assèchement des sols et la déforestation sont responsables pour environ 3% de la montée du niveau des océans, en renvoyant dans l'océan une partie de la quantité d'eau stockée par les continents.

L'érosion des sols, amenant au dépôt de sédiments au fond de l'océan, contribue aussi à la montée du niveau des eaux, sans que le GIEC arrive à le quantifier précisément.

La combustion d'hydrocarbures fossiles, en plus de générer des émissions de gaz à effet de serre, produit de l'eau, qui s'ajoute au cycle de l'eau et finit donc dans l'océan, ce qui contribue à environ 1% de la hausse.

La rétention artificielle d'eau et la construction de barrages est le seul phénomène ayant un impact significatif sur la baisse du niveau des océans. Son effet a compensé environ 10% de la hausse du niveau des océans. A long terme, cet effet sera moindre, du fait de l'envasement des barrages et de la raréfaction des sites propices à la construction de nouveaux barrages.

1.2 La montée des eaux en mer Méditerranée

Nous allons maintenant nous pencher plus précisément sur notre zone de travail, la Méditerranée, et décrire plus en détail l'effet des facteurs décrits ci-dessus [14], [2].

1.2.1 Réchauffement climatique et dilatation thermique

Comme nous l'avons expliqué dans la section précédente, au niveau mondial l'un des principaux moteurs de la montée des eaux est la dilatation thermique des océans. La Méditerranée n'échappe pas à ce phénomène. De plus, en tant que mer semi-fermée, elle ne se comporte pas comme un système isolé et est indirectement impactée via les échanges d'eau avec l'océan Atlantique à travers le détroit de Gibraltar. En effet, les variations du niveau marin dans l'Atlantique peuvent se transmettre partiellement à la Méditerranée, notamment par les ajustements de masse d'eau et de densité. L'augmentation des températures de surface intensifie également les taux d'évaporation, ce qui modifie la salinité des couches supérieures. Ces changements, en influençant la densité et la stratification de la colonne d'eau, peuvent affecter la circulation océanique régionale et, par là même, le niveau moyen de la mer. Ainsi, même si les effets de l'évaporation ou des variations de salinité sur le niveau marin sont indirects et localement variables, ils participent à la dynamique globale du système méditerranéen.

1.2.2 Fonte des glaciers et impact limité en Méditerranée

L'autre facteur majeur que nous avons identifié plus haut est la fonte des glaciers de montagne et des calottes glaciaires.
Cependant, pour plusieurs raisons, cet effet est relativement limité dans notre cas. Même si la Méditerranée n'est pas une mer fermée, elle ne communique qu'en un point avec l'océan

Atlantique, au niveau du détroit de Gibraltar. La fonte des glaciers du Groenland ou de l'Antarctique affecte donc davantage l'océan Atlantique, qui en absorbe directement les effets, là où la Méditerranée ne reçoit que partiellement ces effets. La majorité des glaciers influençant la Méditerranée sont situés dans les Alpes. Même si leur fonte peut avoir un effet local sur certains bassins hydrologiques et les deltas des grands fleuves (comme le Rhône ou le Pô), les glaciers alpins ont une capacité limitée à affecter les masses d'eau océaniques, contrairement à d'autres glaciers et aux calottes glaciaires du Groenland ou de l'Antarctique. De plus, comme nous l'expliquions précédemment, la contribution de ce genre de masse glaciaire va diminuer à long terme du fait de leur disparition.

En résumé, bien que la fonte des glaciers soit un facteur important à l'échelle mondiale, elle n'est pas un moteur principal de la montée des eaux en Méditerranée.

1.2.3 Subsidence et tassements du sol

La subsidence est un enfoncement progressif du sol sous l'effet de facteurs naturels ou anthropiques. Ce phénomène est particulièrement marqué sur les pourtours de la Méditerranée, comme par exemple à Venise, où le niveau de la mer semble monter plus vite qu'ailleurs à cause de l'affaissement du terrain.

Les causes de ce phénomène sont multiples. D'un point de vue géophysique, cela s'explique par la nature des sols et par l'activité tectonique. En effet, dans certaines zones côtières, les sols sont composés de sédiments meubles qui se compactent naturellement sous leur propre poids. De plus, certaines parties du littoral méditerranéen subissent des ajustements lents de la croûte terrestre, contribuant à un affaissement local du terrain. A ces phénomènes naturels s'ajoutent les activités humaines qui viennent aggraver la situation. L'extraction de ressources naturelles comme le gaz et le pétrole ou encore le pompage excessif des nappes phréatiques sont responsables de nombreux affaissements et tassement du sol. A Venise, une étude montre que le pompage des nappes phréatiques à contribué à un enfoncement de la ville de 8 cm entre 1952 et 1969.

Même si ce phénomène ne contribue pas à proprement parler à la montée du niveau des eaux, il participe à la variation relative du niveau de la mer par rapport aux terres. Il est donc important de comprendre ce phénomène dont les conséquences sont similaires à celle d'une élévation du niveau de la mer.

1.2.4 Facteurs tectoniques et déformation de la croûte terrestre

La Méditerranée est une zone géologiquement active, située à la frontière entre plusieurs plaques tectoniques, notamment les plaques africaine et eurasienne. Cette configuration engendre des mouvements lents mais continus de la croûte terrestre, qui peuvent se traduire par des déformations verticales affectant le niveau relatif de la mer.

Ces phénomènes peuvent se manifester de deux manières. D'une part, à travers des processus de déformation lente, comme l'élévation ou l'affaissement progressif de certaines zones côtières sous l'effet de la tectonique des plaques. D'autre part, par des événements sismiques brutaux qui provoquent des ajustements soudains du sol. Un exemple marquant est le séisme de Messine en 1908, qui a entraîné un affaissement local du sol de près de 57 cm, modifiant instantanément le niveau relatif de la mer dans la région.

Bien que ces effets soient généralement localisés, ils peuvent avoir un impact significatif sur l'interprétation des données marégraphiques si on ne les corrige pas. Ils contribuent ainsi

aux variations relatives du niveau marin observées à l'échelle régionale, indépendamment de la montée globale des océans. Il faut donc prendre en compte ces spécificités locales dans l'étude et la compréhension des données.

1.2.5 Oscillations climatiques et refroidissements locaux

L'évolution du niveau marin en Méditerranée n'est pas strictement linéaire : elle est ponctuée de fluctuations liées à des variations climatiques régionales. Par exemple, entre 1950 et 1962, un refroidissement global a temporairement ralenti la montée du niveau de la mer. À Marseille, certaines périodes ont montré des stagnations, voire de légères baisses, en lien avec des changements atmosphériques ou océaniques locaux.

Ces fluctuations s'expliquent par plusieurs facteurs : une baisse des températures atmosphériques limite la dilatation thermique des eaux, tandis que les vents dominants et les phénomènes comme l'oscillation nord-atlantique (NAO) influencent la circulation océanique et la répartition des masses d'eau. Bien que ces variations puissent affecter les tendances sur quelques décennies, elles ne remettent pas en cause la trajectoire générale d'élévation du niveau de la mer.

Encore une fois, il est important de comprendre ces phénomènes régionaux pour exploiter au mieux les données disponibles.

1.2.6 Résumé

Voici les facteurs que nous allons retenir pour la suite de notre travail :

- La température de l'eau, afin de prendre en compte les phénomènes liés au réchauffement climatique et à la dilatation thermique
- La fonte des glaces du Groenland et de l'Antarctique. Bien que leur effet est supposé être inférieur en Méditerranée, cela reste une variable importante. ‘
- La salinité de l'eau et le taux de chlorophylle. Ces variables, bien que secondaires, sont apparues lors de nos recherches et leur ajout permet d'étoffer notre jeu de données.
- Le taux de CO2 dans l'eau. Même si nous ne l'avons pas directement évoqué, l'effet du CO2 sur le changement climatique n'est plus à démontrer, ce qui nous pousse à ajouter cette variable à notre analyse.

Depuis les années 1970, les analyses montrent une élévation continue du niveau marin, sans oscillations marquées, ce qui suggère que l'impact du réchauffement global devient prédominant par rapport aux phénomènes régionaux. C'est la raison pour laquelle nous avons décidé de ne pas ajouter ces éléments à notre analyse.

1.3 Les modèles

L'étude de la montée du niveau des océans repose sur une diversité de modèles permettant d'expliquer et de prédire les fluctuations marines à différentes échelles temporelles et spatiales. Parmi les premiers outils employés, les modèles de régression linéaire ont longtemps servi à capturer les tendances historiques en reliant la montée du niveau des mers à des facteurs tels que l'augmentation des températures moyennes ou la fonte des glaciers (Blanc & Faure, 1990, [2]). Toutefois, ces modèles simplistes pèchent par leur incapacité à prendre en compte les dynamiques complexes des systèmes océaniques et atmosphériques.

Une autre classe de modèles repose sur l'analyse des séries temporelles, avec en tête de file les modèles ARIMA (Auto-Regressive Integrated Moving Average), largement utilisés pour

comprendre l'évolution du niveau des mers (Blanc & Faure, 1990, [2]). Ces modèles permettent d'identifier les tendances et d'extrapoler les niveaux futurs en tenant compte des variations saisonnières et des cycles internes des données.

Le modèle de Holt, un système de lissage exponentiel linéaire, est également appliqué pour traiter les fluctuations du niveau des océans. Ce modèle, en filtrant les variations à court terme, permet d'obtenir des prévisions plus lissées et moins sujettes aux fluctuations erratiques (Blanc & Faure, 1990 [2]). En parallèle, des modèles probabilistes sont utilisés pour intégrer les incertitudes liées aux scénarios de réchauffement climatique. Ces approches s'appuient sur des simulations issues de modèles climatiques globaux (GCM), qui intègrent les interactions entre l'atmosphère, les courants marins, la cryosphère et d'autres paramètres climatiques pour projeter les évolutions à long terme (Sinha et al., 2023, [19]).

Avec l'émergence de nouvelles techniques de traitement des données, les approches basées sur le machine learning ont commencé à être explorées pour améliorer la précision des prédictions. Une des approches les plus classiques demeure la régression multiple, qui permet d'évaluer l'influence de plusieurs variables (température, concentration en CO_2, courant océanique,etc...) sur la montée des eaux (Blanc & Faure, 1990, [2]). Cette technique a été appliquée dans divers travaux pour affiner les tendances observées dans les enregistrements marégraphiques.

Toutefois, les modèles non linéaires comme les forêts aléatoires et les réseaux neuronaux profonds ont montré un potentiel prometteur. Les forêts aléatoires, en combinant plusieurs arbres de décision, réduisent la variance des prévisions et améliorent la robustesse des modèles face à des données bruitées (Sinha et al., 2023, [19]). Les réseaux neuronaux artificiels, quant à eux, sont particulièrement adaptés à la capture des relations complexes entre différents paramètres climatiques. Des études récentes ont exploité ces modèles pour prédire les tendances multidécennales du niveau des mers à partir des données d'altimétrie satellitaire et des ensembles de modèles climatiques (Sinha et al., 2023, [19]).

En complément, des techniques d'apprentissage non supervisé telles que le clustering spectral permettent de segmenter les données spatiales des océans en régions homogènes, facilitant ainsi l'adaptation des modèles à des zones côtières présentant des dynamiques particulières (Sinha et al., 2023, [19]).

Ainsi, si les approches traditionnelles comme les ARIMA et les régressions multiples restent des outils fondamentaux, les nouvelles méthodes basées sur le machine learning ouvrent la voie à des prédictions plus fines et adaptées aux évolutions rapides du climat.

Pour des questions de temps, ne pouvons pas explorer toutes ces options dans notre projet. Par soucis d'interprétabilité, nous avons décidé de concentrer notre travail sur des modèles de régression et de séries temporelles. Ces modèles offrent en effet l'avantage d'être plus facilement interprétables que d'autres méthodes non-linéaires. Ce sont également les méthodes que nous maitrisons le plus, ce qui va donner plus de justesse à notre travail.

Chapitre 2

Analyse des données

Grâce au travail mené dans le chapitre précédent, nous avons désormais des connaissances sur l'état de l'art des projets de prédiction de la montée des eaux. Il s'agit désormais de constituer une base de données pertinente et de l'exploiter pour répondre à notre problématique

2.1 Création de la base de données

Comme nous le présentions dans l'introduction de ce rapport, nous avons eu à constituer nous même notre base de données. Cette étape s'est révélée bien plus fastidieuse que prévue, car même si notre encadrant nous avait averti qu'il ne possédait pas de base ad hoc, nous pensions de prime abord, lui et nous, qu'il nous serait aisé de trouver un jeu de données adapté à notre projet sur internet. Il s'est avéré que ce ne fut pas le cas.

2.1.1 Sélection et origines des données

Dans un premier temps, nous avons cherché à trouver des bases de données contenant toutes les variables que nous avions identifiées lors de la revue de littérature, ou du moins contenant la majorité d'entre elles. Nous avons pour cela éplucher les catalogues de jeux de données de la NASA [15], de la NOAA (National Oceanic and Atmospheric Administration) [16], de l'ESA [8], de Copernicus [4] et du GIEC. Nous nous sommes rapidement rendu compte que ce que nous cherchions n'existait pas, car il y a très peu de bases de données regroupant plusieurs variables à la fois, d'autant plus lorsqu'on restreint la zone géographique à la Méditerranée. Cela vient du fait que les données climatiques sont désormais pour la plupart issues d'observations satellites, et chaque type de satellite mesure une variable précise, ce qui donne des jeux de données univariés.

Nous avons alors décidé de créer nous même notre base de donnée en assemblant à la main des jeux de données univariés sur les variables identifiées plus haut, en choisissant d'utiliser des données mensuelles. Nous nous sommes réparti le travail en cherchant chacun une base pertinente pour une variable précise sur les sites présentés ci-dessus. Pour rappel, ces variables retenues sont les suivantes : niveau de la mer (variable cible), température de l'eau, fonte des glaces en Antarctique, fonte des glaces au Groenland, taux de CO_2 dans l'atmosphère, taux de chlorophylle dans l'eau et salinité de l'eau (variables explicatives). Pour des raisons qui seront exposées plus bas, la recherche et la concaténation des données fut un travail fastidieux, ce qui fait que nous avons décidé de ne pas ajouter d'autres variables dans notre base de données. Même si les variables sélectionnées ont été validées théoriquement, il aurait été intéressant d'ajouter d'autres variables pour étoffer notre travail.

2.1.2 Difficultés techniques

Lors de notre travail pour créer la base de données, nous nous sommes heurtés à différentes difficultés techniques qui sont résumées ci-dessous :

2.1.2.1 Format des données

La plupart des données climatiques disponibles le sont au format NetCDF (Network Common Data Form) qui est un format très pratique pour ce genre de données multidimensionnelles puisque chaque fichier contient à la fois les données (par exemple la température de l'eau mesurée selon le temps, la latitude et la longitude) et les métadonnées associées (nom des variables, unités, dimensions,etc...). Mais même si NetCDF est adapté pour les données climatiques, aucun de nous n'était familier de ce format avant le début du projet, et nous avons donc du prendre le temps d'en comprendre les bases. De plus, il n'est pas aussi aisé de visualiser ce genre de données que celle d'un CSV par exemple, ce qui nous a ralenti ne serait-ce que pour vérifier si un fichier contenait bien des données qui nous intéressent.
Une fois la sélection des jeux de données pertinents effectuée, nous avons également eu à transformer les données au format CSV pour pouvoir faire la modélisation et la prédiction.

2.1.2.2 Concordance temporelle

Une fois le format appréhendé, nous avons fait face à une autre difficulté. Bien que nous ayons trouvé des jeux de données pour chaque variable, les plages temporelles de chacune des bases ne concordent pas totalement. Nous avons des données qui remontent jusqu'à 1982 et d'autres qui ne commencent qu'en 2010. Nous avons bien sûr gardé toute cette plage temporelle dans la base de données, mais nous ne disposons de l'observation de toutes les variables en même temps que de 2010 à 2022, ce qui ne fait que 156 observations exploitables.

2.1.2.3 Concordance géographique

En plus du problème de concordance temporelle, nous avons du faire face à un problème de concordance géographique. Certaines des bases univariées sont proposées avec un maillage très précis sur la Méditerranée, ce qui offre beaucoup de points et de mesure à exploiter, mais d'autres ne proposent pas un maillage aussi précis voir ne proposent qu'une seule valeur moyenne pour l'ensemble de la zone géographique. Cela nous a forcé à moyenner toutes les données avec plus de précision géographique pour se restreindre à une valeur moyenne sur la Méditerranée pour chaque variable.
A ce stade, notre base de données est donc une série temporelle multivariée de dimension 7.

2.1.3 Traitements effectués

Une fois la base de données finalisée, il reste quelques étapes de traitement à effectuer. Nous avons commencé par renommer les variables et formater les données pour avoir une base plus agréable sur laquelle travailler. Par la suite, nous avons remarqué qu'il nous manquait certaines données pour certaines variables sur la période 2010-2022, commune à toutes nos données. Nous avons donc décidé d'interpoler les quelques valeurs manquantes pour compléter notre base de données. Nous avons fait le choix de ne pas interpoler les valeurs manquantes aux bords de nos séries temporelles. Cela aurait permis de gonfler le nombre d'observations exploitables, mais les différentes variables de notre base sont supposées être dépendantes les unes des autres donc cela ne nous semblait pas pertinent de tenter de compléter chaque série sans prendre en compte les liens avec les autres séries.

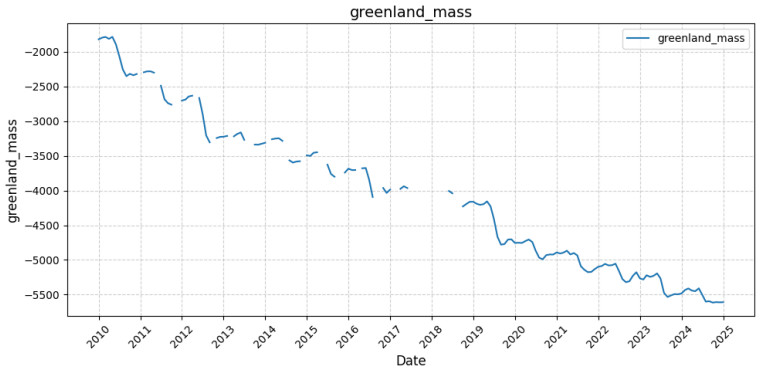

FIGURE 2.1 – Variable masse de glace au Groenland avant l'interpolation

Comme nous avions également remarqué que nos données avaient l'air saisonnalisées, le traitement suivant a été de désaisonnaliser les variables qui ne l'étaient pas déjà. Pour cela nous avons appliqué la transformation suivante : $X_t - X_{t-12}$

Pour la plupart des méthodes que nous allons utiliser par la suite, il est préférable voir impératif que les données soient standardisées afin d'éviter des effets liés à l'ordre de grandeur des variables de la base. Nous avons fait le choix d'ajouter la standardisation dans la pipeline de chaque régression plutôt que de travailler sur une base standardisée. Il en va de même pour les différenciations d'ordre 1 dans les modèles de séries temporelles.

2.1.4 Présentation de la base de données

Voici la base de données sur laquelle nous avons mené le projet. La description des variables est disponible dans la table ci-dessous. Suite à l'étape de désaisonnalisation, notre base contient 144 observations complètes, une par mois, de janvier 2011 à décembre 2022. Ce sont ces observations qui serviront pour l'analyse multivariée. Pour les statistiques descriptives et l'analyse univariée, on se sert de la série complète pour chaque variable.

Variable	**Nom**	**Unité**	**Description**
Niveau de la mer	`sea_level`	cm	Variation du niveau de la mer par rapport à une valeur de référence
Température de surface	`sea_temperature`	°C	Variation de la température de surface par rapport à une valeur de référence
Taux de CO_2 atmosphérique	`CO2`	mmol/mol	Fraction molaire de CO_2 dans l'air sec
Taux de chlorophylle dans l'eau	`chlorophylle`	mg/m^3	Concentration de chlorophylle-a dans la mer Méditerranée
Salinité	`sea_salinity`	psu (g/L)	Concentration de sel dans l'eau de mer
Masse de glace en Antarctique	`antarctica_mass`	Giga Tonne	Variation de la masse de glace de l'Antarctique par rapport à une valeur de référence
Masse de glace au Groenland	`greenland_mass`	Giga Tonne	Variation de la masse de glace du Groenland par rapport à une valeur de référence

TABLE 2.1 – Présentation des variables de la base de données

Un extrait de la base de données est également disponible en annexe.

2.2 Statistiques descriptives

Dans cette section nous allons effectuer quelques statistiques descriptives sur notre base de données, essentiellement pour vérifier que les valeurs de la base sont cohérentes avec les connaissances physiques sur l'évolution théoriques de ces variables. On regarde ici les variables non désaisonnalisées pour avoir une meilleure compréhension des phénomènes sous-jacents.

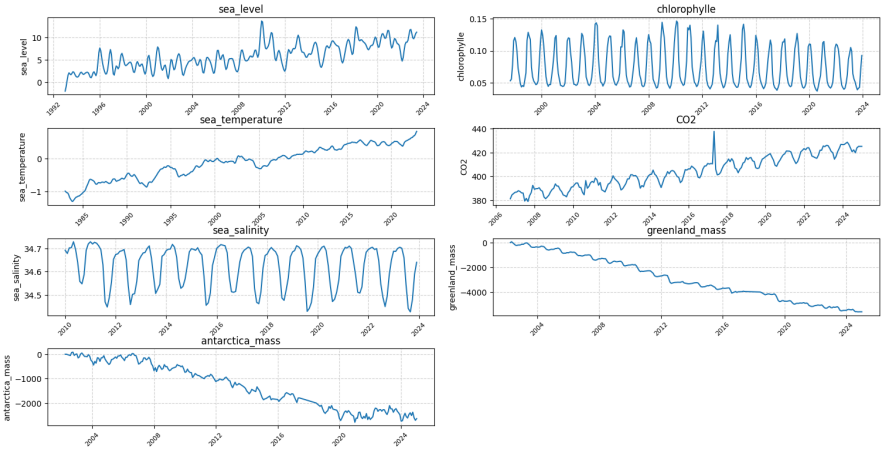

FIGURE 2.2 – Evolution des variables de la base de données

2.2.1 Niveau de la mer

Pour rappel, la variable de notre base codant le niveau de la mer [18] mesure la variation du niveau de la mer par rapport à une valeur de référence (apparemment fixée en 1993 d'après le graphique). Le niveau de la mer présente une tendance clairement croissante sur l'ensemble de la période observée (1992–2023), avec des dynamiques irrégulières qui n'ont pas l'air de s'apparenter à une saisonnalité. L'élévation moyenne est d'environ 10 cm sur la période observée. L'évolution de cette série confirme l'expression 'montées des eaux' et est pertinente en tant que variable cible de notre projet de prédiction.

2.2.2 Température de surface

La variable de notre base qui traite de la température de surface de la mer [18], et qui nous sert de proxy de la dilatation thermique, mesure la variation de température par rapport à une valeur de référence. La température de surface de la mer suit également une nette tendance à la hausse, avec une élévation moyenne d'environ 2 degrés sur les 40 dernières années. Cette élévation reflète le réchauffement climatique en Méditerranée et explique une partie de la dilatation thermique des océans ainsi que la fonte des glaces, deux phénomènes à l'origine de la montée du niveau marin.

2.2.3 Masses de glaces au Groenland et en Antarctique

On traite ces 2 variables conjointement car elles reflètent sensiblement le même phénomène à deux endroits différents. Les variables de notre base mesure la différence de masse de glace par rapport à une valeur de référence (fixée en avril 2002 d'après la documentation)[15]. Au Groenland comme en Antarctique, on observe une baisse marquée et continue sur la période. Le Groenland a perdu près de 6000 Gt de glace depuis 2002, tandis que la perte est d'environ 2000 à 3000 Gt de glace pour l'Antarctique.

2.2.4 Chlorophylle

Notre variable mesure la concentration de chlorophylle dans la mer Méditerranée [5]. Elle suit un cycle saisonnier très marqué, avec des pics récurrents correspondant probablement aux périodes de floraison du phytoplancton (généralement au printemps). Ce comportement justifie une désaisonnalisation avant toute analyse statistique ou modélisation.

2.2.5 CO_2

Notre variable mesure le taux de CO_2 atmosphérique [17]. On remarque une croissance progressive et régulière (de 380 à 420 mmol/mol entre 2007 et 2024), en ligne avec les tendances mondiales d'augmentation des émissions. Malgré une valeur probablement aberrante en 2017, la dynamique globale reste très stable. Un effet saisonnier est également visible, ce qui justifie là aussi le besoin d'une désaisonnalisation.

2.2.6 Salinité

Notre variable mesure la concentration de sel dans l'eau de mer [4]. La salinité montre une forte saisonnalité, vraisemblablement liée aux phénomènes d'évaporation et d'apports d'eau douce (précipitations, fleuves). Aucune tendance longue nette n'apparaît clairement sur la période disponible (2010–2023), ce qui justifie également une désaisonnalisation préalable. On peut également noter que les variations sont très faibles (entre 34,5 et 34,7 sur la période).

2.3 Justification empirique des variables : Premières régressions linéaires

Le but de cette section est d'évaluer empiriquement l'impact de chaque variable sur le niveau de la mer, afin de valider les résultats théoriques évoqués plus haut.
Pour cela, on effectue une régression linéaire simple de chaque variable explicative sur la variable cible. On utilise la méthode des moindres carrés pour estimer le coefficient de la variable explicative.

2.3.1 Effet de la température de surface sur le niveau de la mer

On effectue la régression linéaire du niveau de la mer sur la température de surface. On obtient alors les résultats suivants.

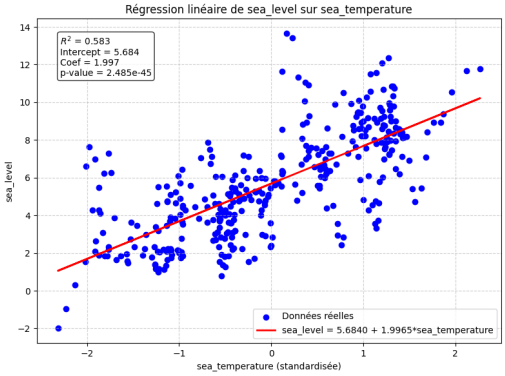

FIGURE 2.3 – Régression linéaire du niveau de la mer sur la température de surface

Cette régression montre un effet significatif de la température de surface sur le niveau de la mer. L'effet semble important et la relation linéaire semble crédible au vu du nuage de point. Le coefficient associé à la température de surface de la mer est de $1,997$. La p-valeur indique que le coefficient associé à la température dans la régression étudiée est significatif aux seuils usuels. Le $R^2 = 0,583$ est satisfaisant.
Il semblerait qu'une augmentation d'un écart-type de la température de surface $(0,12C)$ soit associée à une hausse d'environ deux centimètres du niveau de la mer. Le sens de variation est cohérent avec ce à quoi on pourrait s'attendre.

2.3.2 Effet de la fonte des glaces au Groenland et en Antarctique sur le niveau de la mer

On effectue les régressions linéaires simples du niveau de la mer sur la masse de glace au Groenland (resp. en Antarctique). On obtient alors les résultats suivants.

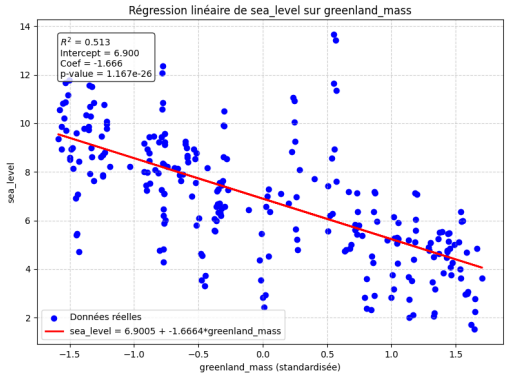

FIGURE 2.4 – Régression linéaire du niveau de la mer sur la masse de glace au Groenland

Cette régression montre un effet significatif de la variation de la masse de glace au Groenland sur le niveau de la mer. La relation linéaire semble crédible au vu du nuage de point, même si on observe quelques point dispersés. Le coefficient associé à la masse de glace au Groenland est de $-1,66$. La p-valeur indique que le coefficient associé à la température dans la régression étudiée est significatif aux seuils usuels. Le $R^2 = 0,513$ est satisfaisant.

Il semblerait que la fonte d'un écart-type de glace au Groenland $(815Gt)$ soit associée à une augmentation d'environ $1,66$ centimètres du niveau de la mer. Là aussi le sens de variation est cohérent.

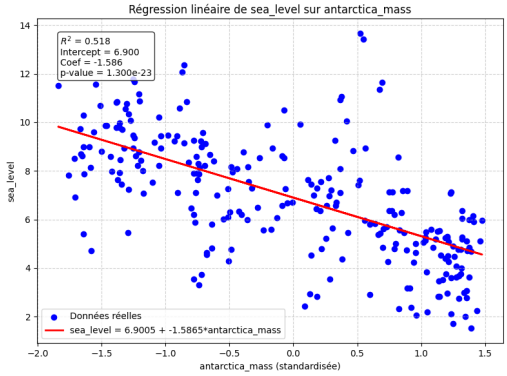

FIGURE 2.5 – Régression linéaire du niveau de la mer sur la masse de glace en Antarctique

On observe des résultats très similaires pour la masse de glace en Antarctique.
Cette régression montre un effet significatif de la variation de la masse de glace en Antarctique sur le niveau de la mer. La relation linéaire semble crédible au vu du nuage de point, même si on observe quelques point dispersés. Le coefficient associé à la masse de glace en Antarctique est de $-1,58$. La p-valeur indique que le coefficient associé à la température dans la régression étudiée est significatif aux seuils usuels. Le $R^2 = 0,518$ est satisfaisant.

Il semblerait que la fonte d'un écart-type de glace en Antarctique ($537Gt$) soit associée à une augmentation d'environ $1,58$ centimètres du niveau de la mer. Comme pour le Groenland, le sens de variation est cohérent.

2.3.3 Effet de la chlorophylle sur le niveau de la mer

On effectue la régression linéaire du niveau de la mer sur le taux de chlorophylle-a dans l'eau. On obtient alors les résultats suivants.

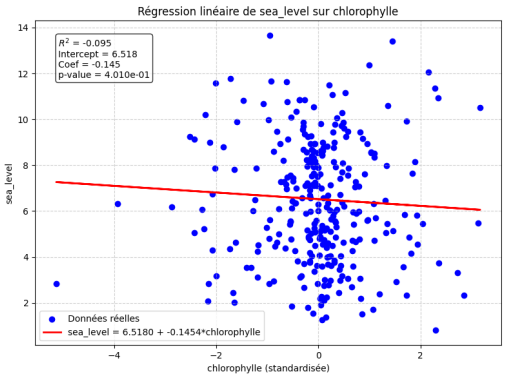

FIGURE 2.6 – Régression linéaire du niveau de la mer sur le taux de chlorophylle

Cette régression montre un résultat difficilement interprétable au vu du R^2 négatif. La relation semble clairement ne pas être linéaire au vu du nuage de point. Pour rappel, un R^2 négatif signifie ici qu'il vaut mieux prédire le niveau de la mer par sa moyenne qu'en utilisant la chlorophylle.

Il est toutefois intéressant de garder cette variable car la chlorophylle-a est utilisée comme indicateur du niveau d'eutrophisation, car elle reflète la concentration d'algues phytoplanctoniques, qui croissent fortement lors des apports excessifs en nutriments.
L'eutrophisation est un enrichissement excessif des eaux en nutriments (azote et phosphore). C'est aujourd'hui l'une des principales causes de dégradation de la qualité de l'eau, notamment dans les zones côtières.
Une concentration élevée de chlorophylle-a indique une forte activité biologique, souvent liée à des blooms algaux (c'est-à-dire une prolifération d'algues), eux-mêmes responsables de l'appauvrissement en oxygène.
Ce manque d'oxygène altère la vie marine, y compris celle des organismes qui participent au cycle du carbone et à la sédimentation. Ces deux facteurs sont indirectement liés à la stabilité des fonds marins et donc à l'évolution locale du niveau de la mer.
L'eutrophisation peut donc modifier la structure des écosystèmes côtiers, affecter les zones humides, les zones tampons contre l'élévation du niveau marin, et provoquer la dégradation des herbiers marins qui jouent un rôle dans la stabilisation du littoral.
À long terme, ces perturbations peuvent réduire la résilience des zones côtières face à la montée

des eaux, notamment en fragilisant les écosystèmes protecteurs comme les mangroves ou les marais salants.

Ainsi, en tant qu'indicateur de l'eutrophisation et des dynamiques biologiques côtières, la chlorophylle-a peut fournir des informations utiles pour anticiper les mécanismes écologiques susceptibles d'influencer localement l'élévation du niveau de la mer.

2.3.4 Effet du taux de CO_2 sur le niveau de la mer

On effectue la régression du niveau de la mer sur le taux de CO_2 atmosphérique. On obtient alors les résultats suivants.

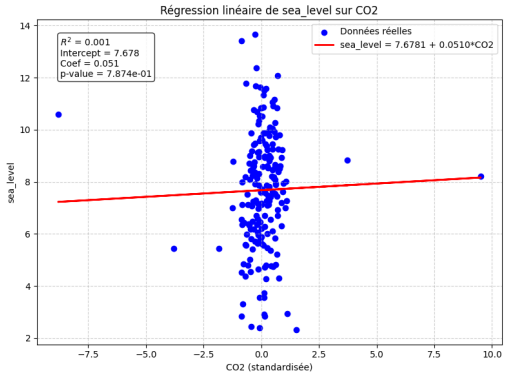

FIGURE 2.7 – Régression linéaire du niveau de la mer sur le taux de CO_2

Les résultats montrent que le modèle linéaire est très peu crédible. Le R^2 est très proche de 0.

Toutefois, comme la chlorophylle il est intéressant de garder le CO2 comme proxy de l'acidification des océans.

En effet, l'acidification des océans résulte directement de l'absorption par les océans d'une part importante du CO_2 émis par les activités humaines. Environ 25 % du CO_2 atmosphérique est absorbé par les mers, où il réagit chimiquement pour former de l'acide carbonique, entraînant une baisse progressive du pH de l'eau. Ce phénomène perturbe l'équilibre de la chimie des carbonates, indispensable à la formation des structures calcaires comme les coquilles, les squelettes.
Cette acidification représente une menace majeure pour les organismes marins calcifiants comme les coraux, le phytoplancton ou les mollusques, affectant ainsi la base de la chaîne alimentaire et le fonctionnement des écosystèmes côtiers. En particulier, les récifs coralliens, essentiels pour la biodiversité et les services écosystémiques, sont fragilisés par l'acidification.
Cette acidification réduit la capacité des récifs coralliens à protéger les côtes contre l'érosion et les tempêtes, exposant davantage les littoraux à la montée du niveau marin.
Le phytoplancton affecté joue aussi un rôle dans le cycle du carbone et la séquestration naturelle du CO_2, influençant indirectement le réchauffement climatique et donc la dilatation thermique des océans, l'un des principaux facteurs de la montée du niveau de la mer.

2.3.5 Effet de la salinité sur le niveau de la mer

On effectue la régression linéaire du niveau de la mer sur la salinité. On obtient alors les résultats suivants.

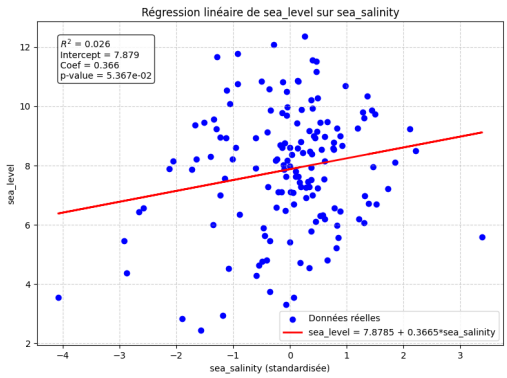

FIGURE 2.8 – Régression linéaire du niveau de la mer sur la salinité

Cette régression montre un résultat plus mitigé. Le nuage de point montre une régression linéaire peu crédible. Le modèle n'explique pas bien le lien entre la salinité et le niveau de la mer comme confirmé par le R^2 faible. Cela n'est pas étonnant au vu des recherches sur la salinité plaçant ce facteur comme un facteur secondaire. Nous gardons quand même ce facteur pour la suite de notre travail, car il se pourrait qu'il joue un rôle dans la montée des eaux via des mécanismes plus complexes et non linéaires, que nous allons explorer plus tard.

2.4 Une première tentative de modélisation : La régression multiple

Après avoir étudié séparément la relation entre le niveau de la mer et chaque variable explicative, nous proposons ici une première régression multiple afin de mieux comprendre l'effet conjugué de l'ensemble des variables. Cette approche permet de mieux isoler l'effet propre de chaque facteur tout en tenant compte de la présence possible de corrélations entre les variables.

L'objectif est principalement exploratoire : il s'agit d'évaluer si les tendances observées dans les régressions simples se maintiennent lorsque toutes les variables sont considérées simultanément. L'analyse qui suit porte donc sur l'interprétation des coefficients estimés, leur signe, leur significativité, et leur cohérence avec les résultats précédents.

OLS Regression Results

Dep. Variable:	sea_level	R-squared:	0.292
Model:	OLS	Adj. R-squared:	0.261
Method:	Least Squares	F-statistic:	9.437
Date:	Fri, 02 May 2025	Prob (F-statistic):	1.15e-08
Time:	13:51:11	Log-Likelihood:	-281.01
No. Observations:	144	AIC:	576.0
Df Residuals:	137	BIC:	596.8
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	7.7693	0.146	53.391	0.000	7.482	8.057
sea_temperature	-0.0250	0.240	-0.104	0.917	-0.499	0.449
greenland_mass	-1.2745	0.459	-2.777	0.006	-2.182	-0.367
antarctica_mass	0.2297	0.414	0.555	0.580	-0.589	1.049
chlorophylle	0.0109	0.149	0.073	0.942	-0.284	0.306
CO2	-0.0557	0.149	-0.375	0.708	-0.349	0.238
sea_salinity	0.2421	0.154	1.570	0.119	-0.063	0.547

Omnibus:	1.723	Durbin-Watson:	0.264
Prob(Omnibus):	0.422	Jarque-Bera (JB):	1.373
Skew:	-0.230	Prob(JB):	0.503
Kurtosis:	3.135	Cond. No.	6.73

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

FIGURE 2.9 – Sortie de la régression linéaire multiple

2.4.1 Analyse des résultats

Les résultats de cette régression sont assez étonnants. D'après la sortie ci-contre, seul le coefficient lié à fonte des glaces au Groenland est significatif aux seuils habituels. On peut d'ailleurs remarquer que le signe et la valeur sont semblables à ce qu'on avait pour la régression simple du niveau de la mer sur la masse de glace au Groenland.

Pour ce qui est des autres coefficients, ils ne sont pas significatifs aux seuils habituels, il faut donc les analyser avec prudence.

Les coefficients de $sea\_temperature$, de $antarctica\_mass$ et de CO_2 ont un signe contre intuitif, on s'attendrait plutôt au signe opposé d'après les connaissances physiques que nous avons de la montée des océans. Mais comme ces coefficients ne sont pas significatifs, cela pourrait simplement être dû au hasard.

Le coefficient de la chlorophylle est très proche de 0 et fortement non significatif (p-value = 0.942), ce qui suggère qu'il n'y a aucune relation détectable avec le niveau de la mer, même si sur le plan écologique des liens indirects sont envisageables.

Le coefficient associé à la salinité est proche d'un seuil de significativité acceptable, Cela pourrait refléter une tendance où une augmentation de la salinité est associée à une élévation du niveau marin, peut-être en lien avec les changements de densité de l'eau ou les courants océaniques. Ce résultat mériterait exploration dans un modèle plus robuste ou avec plus de données.

De plus, ces résultats laissent penser qu'il pourrait y avoir de la colinéarité entre nos variables. En effet, bien que l'effet de la température ne soit pas statistiquement significatif, il pourrait être masqué par une forte corrélation avec les variables représentant la fonte des glaces ($antarctica\_mass$ et $greenland\_mass$), qui sont toutes liées au réchauffement climatique.

2.4.2 Limites de la régression multiple

Comme nous le présentions plus haut, l'objectif de cette modélisation est avant tout exploratoire. En effet, la régression multiple passe sous silence la plupart des spécificités structurelles de notre jeu de données. Sur le plan statistique, la présence de colinéarité entre certaines variables affaiblit la stabilité des coefficients estimés. Surtout, la dimension temporelle de nos séries n'est pas modélisée, ce qui expose la régression au risque de régression fallacieuse si les variables sont non stationnaires et non co-intégrées.
En outre, d'un point de vue physique, les phénomènes qui gouvernent la montée du niveau de la mer sont complexes et possiblement non linéaires — une simple relation linéaire ne saurait donc suffire à les modéliser avec fidélité.

Ce sont ces problèmes que nous allons désormais tenter de résoudre pour arriver à un modèle permettant une prévision robuste du niveau de la mer.

Chapitre 3

Modélisation & prédiction

À la lumière des limites identifiées dans le chapitre précédent — notamment la possible colinéarité entre variables, la non-linéarité des relations physiques sous-jacentes, et l'absence de prise en compte de la dynamique temporelle — nous explorons ici des méthodes de modélisation plus avancées, dans le double objectif de mieux comprendre les mécanismes à l'œuvre et d'obtenir des prévisions fiables de l'évolution du niveau de la mer.

3.1 Choix des modèles

Ce chapitre est structuré en deux temps. Dans un premier temps, nous enrichissons le cadre de la régression en introduisant des effets non linéaires et des régularisations pour mieux gérer la colinéarité. Dans un second temps, nous passons à des modèles dynamiques adaptés aux séries temporelles multivariées, en particulier les modèles VAR et VECM, qui nous permettent d'intégrer la dépendance temporelle et les relations de long terme entre variables.

Avant de présenter ces modèles étendus, il est essentiel de s'attarder sur les limites structurelles identifiées plus haut. Ces limites concernent notamment la présence de colinéarité entre variables explicatives, l'existence possible de relations non linéaires mal capturées par les modèles linéaires standards, et la non-stationnarité des séries temporelles impliquant des dynamiques de long terme. Les sections qui suivent détaillent ces problématiques à travers des visualisations, des justifications théoriques et des tests statistiques, afin de mieux motiver le choix des modèles plus avancés présentés ensuite.

3.1.1 Colinéarité entre les variables

D'après la nature de nos données, on peut supposer une colinéarité entre nos données, on va donc tester si c'est le cas.

On peut voir, comme on s'y attendait, que certaines de nos variables sont fortement corrélées entre elles.
La corrélation la plus évidente est celle entre la fonte des glaces au Groenland et en Antarctique (corrélation de $0,93$ entre $antarctica\_mass$ et $greenland\_mass$). Cela vient sans doute du fait que la fonte des glaces est essentiellement due à la hausse des températures sur le globe. Même si d'autres facteurs entrent en compte et que la température n'est pas la même au Groenland qu'en Antarctique, il y a donc une forte corrélation entre les 2 variables. Cette explication physique permet aussi de comprendre la forte corrélation négative entre la température de surface de la mer et les masses de glace (corrélation de $-0,78$ avec $greenland\_mass$, de $-0,72$ avec $antarctica\_mass$). Ces variables sont toutes fortement corrélées à la température de l'atmosphère terrestre, ce qui explique cette matrice de corrélation.

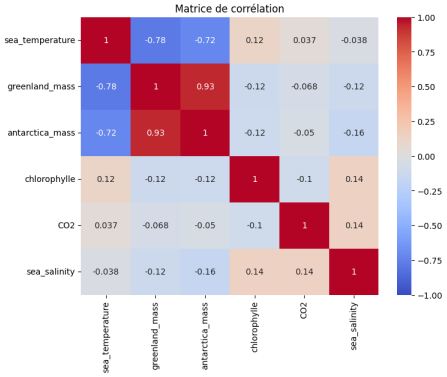

FIGURE 3.1 – Matrice de corrélation

Même si le reste de la matrice ne met pas en avant d'autres phénomènes de colinéarité flagrants, nous allons désormais prendre en compte la colinéarité dans nos régressions en utilisant une régularisation Ridge. Cela permettra d'avoir une meilleure stabilité des coefficients estimés.

3.1.2 Non-linéarité dans les mécanismes de la montée des eaux

Dans cette partie, on cherche à rappeler plus précisément la théorie derrière les variables sélectionnées. Rappelons que d'après le copernic service maritime [6], sur la période 1993-2022, le niveau moyen de la mer s'est élevé en moyenne de 3,3 ± 0,3 mm par an. Un total de 9,57 cm sur cette période. Cette élévation est principalement due comme mentionné plus haut à la dilatation thermique des océans et la fonte des glaces.
Cherchons à comprendre les équations physiques liées à ces phénomènes.
Pour cela on se propose d'expliquer plus finement les liens non linéaires évoqués en revue de littérature quant à la densité de l'eau [1]. On sait que la densité de l'eau qui est exprimée en g/cm^3 varie fortement en fonction de la température, de la pression mais aussi de la salinité. On retrouve donc ici les facteurs choisis. Toutefois cette relation est complexe et non linéaire et suit le graphique suivant. Le diagramme Température–Salinité ci-dessous sert à comprendre comment la densité de l'eau de mer change selon sa température (en °C) et sa salinité (quantité de sel, exprimée en PSU). Les lignes courbes tracées sur le graphique s'appellent isopycnes. Ce sont des lignes qui relient tous les points ayant la même densité de l'eau. Par exemple, une eau à haute salinité et chaude peut avoir la même densité qu'une eau froide mais moins salée. Ces deux eaux se retrouveront sur la même isopycne. On voit donc ici que la dépendance entre ces trois valeurs n'est pas constante.

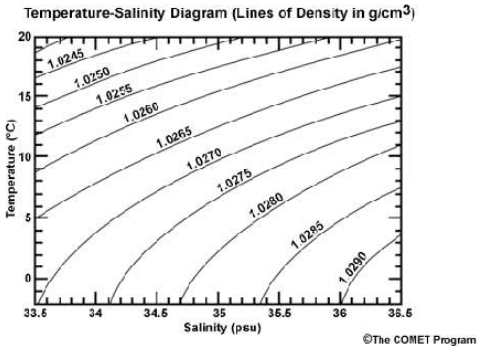

FIGURE 3.2 – Diagramme Température–Salinité avec isopycnes représentant les variations de densité de l'eau de mer en fonction de la température et de la salinité

Il est donc important de voir ici que densité, température et salinité dont fortement liées et ce de manière non-linéaire. On peut donc supposer que des modèles non-linéaires pourront capter plus précisément le lien entre ces variables.

On propose maintenant de présenter un modèle physique et mathématique mettant en jeu les deux facteurs principaux identifiés afin de mieux comprendre les relations que nous sommes supposés retrouver [1].

Ce modèle donne le résultat suivant (les calculs, non essentiels à la compréhension, sont développés en annexe) :

$$h = \frac{m_I}{4\pi R^2 \rho_w f}. \tag{3.1}$$

Où h représente l'élévation du niveau de la mer due à la fonte des glaces, m_I la masse de glace fondue, R le rayon de la terre, ρ_w la densité de l'eau et f la fraction de la surface terrestre couverte par les océans.

La température joue un rôle crucial en affectant la densité de l'eau de mer : une hausse de température entraîne une diminution de la densité, ce qui provoque une expansion du volume et donc une élévation du niveau marin. Ce lien est non linéaire, car la densité varie de manière complexe avec la température. La fonte des glaces continentales, quant à elle, contribue linéairement dans ce modèle à la hausse du niveau marin en ajoutant directement du volume d'eau dans les océans. C'est un résultat que nous retrouvons avec nos régressions linéaires. La salinité influence indirectement cette dynamique en modulant la densité de l'eau : une eau plus salée est plus dense et se dilate moins. D'autres facteurs comme le taux de chlorophylle ou le CO_2 atmosphérique interviennent de manière indirecte : la chlorophylle est corrélée à certaines conditions climatiques mais sans effet direct sur h, tandis que le CO_2, principal moteur du réchauffement climatique, agit en amont en amplifiant la température et la fonte des glaces.

3.1.3 Stationnarité des séries et relations de cointégration

L'autre point important que nous avons passé sous silence dans les premières modélisations est la dynamique temporelle de nos données. Il est raisonnable de penser qu'il y a une dépendance temporelle entre les valeurs de nos variables, ce qui peut être le signe d'une non stationnarité. Cette hypothèse est renforcée par l'observation de tendances visibles dans plusieurs séries.

Nous avons donc mené des tests de stationnarité sur nos séries pour savoir si certaines séries sont non stationnaires. Nous avons opté pour un test ADF couplé à un test KPSS pour chaque variable, avec un niveau fixé à 5%. Pour rappel, les hypothèses H_0 des deux tests sont opposées. Pour le test ADF, l'hypothèse nulle correspond à la présence de racine unitaire, on suppose donc la série stationnaire si la p-valeur est inférieure à 5%. Pour le test KPSS, l'hypothèse nulle est la stationnarité de la série, on suppose donc la série stationnaire si la p-valeur est supérieure à 5%.

	Variable	ADF p-value	ADF conclusion	KPSS p-value	KPSS conclusion
0	sea_level	0.0462	Stationnaire	0.01	Non stationnaire
1	sea_temperature	0.9952	Non stationnaire	0.01	Non stationnaire
2	greenland_mass	0.7528	Non stationnaire	0.01	Non stationnaire
3	antarctica_mass	0.4084	Non stationnaire	0.01	Non stationnaire
4	chlorophylle	0.0002	Stationnaire	0.10	Stationnaire
5	CO2	0.0006	Stationnaire	0.10	Stationnaire
6	sea_salinity	0.0127	Stationnaire	0.10	Stationnaire

FIGURE 3.3 – Résultats des tests de stationnarité

Les résultats ci-dessus font état d'une non stationnarité pour les variables suivantes : *sea_level*, *sea_temperature*, *greenland_mass* et *antarctica_mass* (pour *sea_level*, les tests donnent des résultats différents, nous avons décidé de donner plus de crédit au test KPSS, réputé plus adapté, d'autant que la p-valeur du test ADF est proche du seuil de 5%). Ces résultats justifient le fait de considérer des modèles tels que VAR ou VECM pour la modélisation.

Passons maintenant aux tests de relation de cointégration entre les variables non stationnaires. Les résultats permettront de choisir entre un modèle VAR (pas de relation de cointégration) ou VECM (présence de relations de cointégration).

Nous avons commencé par vérifier que les séries différenciées sont bien stationnaires. Nous avons à nouveau conduit un double de test ADF et KPSS. Les résultats disponibles en annexe nous font conclure à la stationnarité des séries différenciées. On parle alors de séries I(1) (intégrées d'ordre 1) pour les séries qui sont stationnaires seulement après une différenciation.

Nous avons ensuite testé la cointégration de deux manières différentes : un test d'Engle-Granger (test de stationnarité ADF sur les résidus de la régression multiple) et un test de Johansen (test du nombre de relations de cointégration). Les deux tests sont conduits avec un niveau de 5%.

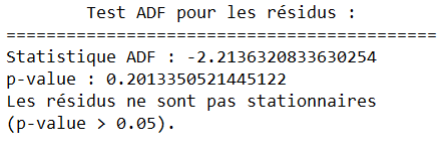

FIGURE 3.4 – Résultats du test d'Engle-Granger

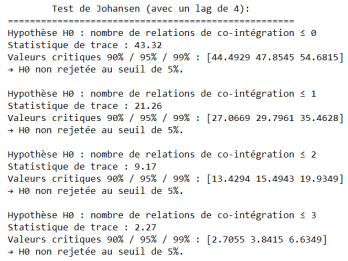

FIGURE 3.5 – Résultats du test de Johansen (lag 4)

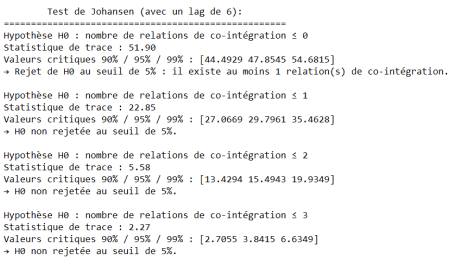

FIGURE 3.6 – Résultats du test de Johansen (lag 6)

Les tests fournissent des résultats différents. Le test d'Engle-Granger nous dit que les résidus de la régression ne sont pas stationnaires, ce qui signifie qu'il n'y a pas de relation de cointégration et qu'un modèle VAR serait le plus adapté. Mais les tests de Johansen ne donnent pas forcément les mêmes résultats. Selon la valeur de retard choisie, on obtient ou non qu'il existe au moins une relation de cointégration. On montre ici les résultats pour les retards 4 et 6, jugés optimaux par les critères AIC ou BIC (cf modèle VECM). Le test de Johansen avec un retard de 6 périodes nous dit qu'il existe une relation de cointégration, ce qui signifie qu'un modèle VECM serait le plus adapté.

À la vue de ces résultats, nous avons décidé de commencer par envisager le modèle VAR, et nous nous pencherons sur le modèle VECM par la suite.

3.2 Régressions polynomiales

L'objectif de cette section est de comprendre comment une régression polynomiale peut être utilisée pour prédire le niveau de la mer en tenant compte de nos variables explicatives. Le choix de la régression polynomiale vient du besoin de modéliser des relations non linéaires entre ces variables et la variable cible. En effet, nous avons vu qu'une régression linéaire classique

pouvait ne pas saisir toutes les interactions complexes qui existent entre ces variables.

Dans une régression polynomiale, l'idée est d'ajouter des termes de puissance supérieure aux variables explicatives pour capturer les comportements non linéaires. Par exemple, si une variable X est utilisée, au lieu de simplement l'inclure dans le modèle comme un terme linéaire (βX), on ajoute également des termes quadratiques (βX^2) ou des interactions croisées entre différentes variables explicatives ($\beta X_1 X_2$).

Dans notre cas, nous avons choisi de limiter le degré du polynôme à 2, ce qui signifie que nous allons inclure non seulement les termes linéaires des variables explicatives, mais aussi leurs carrés et leurs produits croisés. Le choix d'un degré limité à 2 a été motivé théoriquement par les risques de sur-apprentissage liés à l'explosion du nombre de paramètres, et soutenu empiriquement par les résultats d'une validation croisée, qui ont montré qu'un degré plus élevé n'apportait pas de gains significatifs et augmentait le risque de sur-apprentissage, ce qui peut nuire à la performance du modèle sur des données non observées.

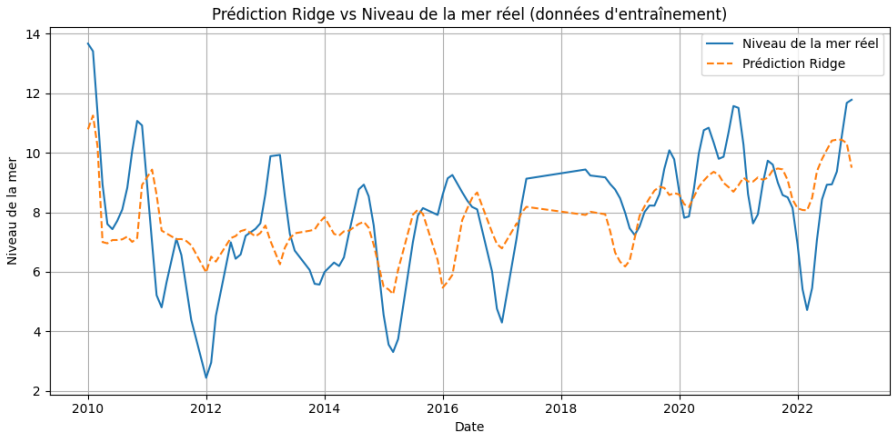

FIGURE 3.7 – Prédiction sur les données d'entraînement avec degré 3

On montre ici le résultat d'une régression polynomiale de degré 3. Par rapport à un degré 2, l'erreur sur les données d'entraînement est plus faible, par contre le modèle se généralise mal sur les données de test.

Comme nous le présentions plus haut, une régularisation Ridge a été appliquée sur ce modèle. Elle est d'autant plus intéressante dans la régression polynomiale car la colinéarité se propage dans les termes d'ordre 2. Elle permet donc de stabiliser le modèle en réduisant la variance, tout en préservant les coefficients pertinents.

Concernant la division des données en ensemble d'entraînement et ensemble de test, nous avons opté pour un découpage chronologique des données, car les séries temporelles présentent une forte dépendance entre les observations successives. L'ensemble d'entraînement représente 90% des données, et l'ensemble de test, 10%. Ce choix permet d'entraîner le modèle sur une quantité suffisante d'informations, ce qui est particulièrement pertinent dans un contexte temporel où la structure des données doit être bien apprise. Cette séparation est essentielle pour évaluer la capacité du modèle à prédire de nouvelles données sans que celles-ci ne contaminent l'entraînement.

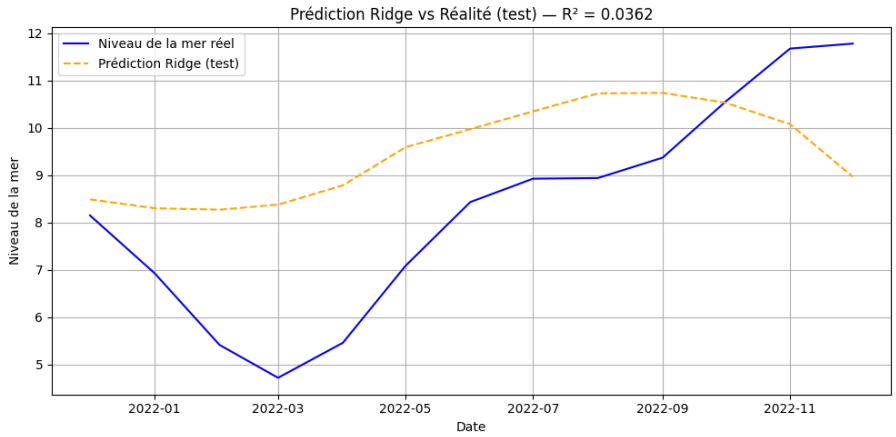

FIGURE 3.8 – Régression polynomiale Ridge - prédiction sur les données de test

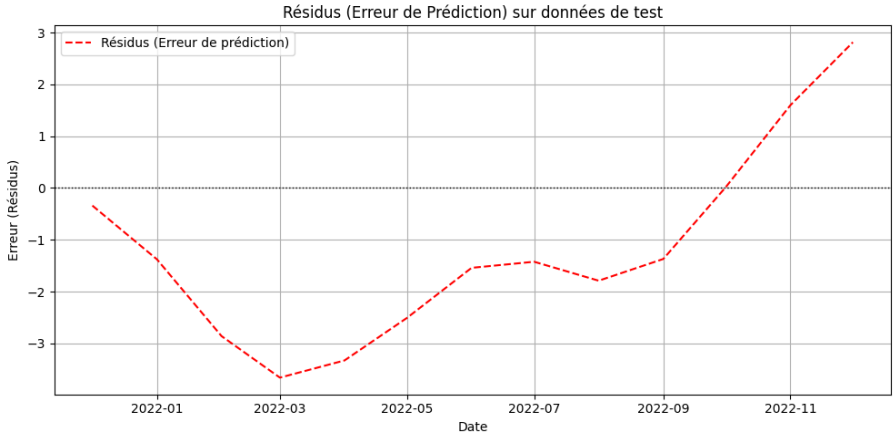

FIGURE 3.9 – Régression polynomiale Ridge - résidus sur les données de test

En observant les résultats obtenus sur les données de test, on constate que le modèle parvient à capturer les tendances générales du niveau de la mer, mais échoue à prédire correctement les valeurs absolues. Cette inadéquation entre les prédictions et les observations réelles se traduit par des résidus non négligeables, et un décalage systématique entre les courbes.

Pourtant, les données ont été standardisées en amont, ce qui rend ce décalage d'autant plus surprenant. Le R^2, bien que positif, reste faible, ce qui reflète une capacité prédictive limitée. Par ailleurs, l'erreur quadratique moyenne (MSE) obtenue sur le jeu de test est de 4.6851, ce qui confirme cette faiblesse de prédiction.

Des pistes d'amélioration sont envisageables, notamment l'introduction de lag features, c'est-à-dire de variables décrivant les valeurs passées du niveau de la mer ou d'autres variables explicatives. Ce type de transformation permettrait au modèle de mieux intégrer la dépendance temporelle inhérente aux séries chronologiques, et donc d'améliorer la précision des prédictions.

Au vu des résultats des tests de stationnarité et de cointégration, nous avons décidé de préférer une approche par les séries temporelles plutôt que d'améliorer ces modèles de régression. Cela va nous permettre une meilleure prise en compte des dynamiques de nos jeux de données, ce qui nous semble pertinent dans un but prédictif.

3.3 Modèle VAR

En nous basant sur les résultats des tests de stationnarité et du test de cointégration d'Engle-Granger, nous avons décidé d'estimer un modèle vectoriel auto-régressif (VAR). Pour ce faire, nous avons dû stationnariser les variables identifiées comme ne l'étant pas, c'est une étape nécessaire pour des prévisions valides et non biaisées.
En se basant sur les critères AIC et HQIC qui garantissent un bon compromis entre parcimonie et performance prédictive, nous avons sélectionné un modèle VAR d'ordre 2, ce qui signifie qu'on utilise les valeurs des variables en $t-1$ et en $t-2$ pour prédire les variables en t.

Le modèle estimé s'écrit sous la forme :

$$\Delta y_t = c + A_1 \Delta y_{t-1} + A_2 \Delta y_{t-2} + \varepsilon_t$$

où Δy_t désigne le vecteur des variables différenciées au temps t (Confère l'annexe pour la spécification théorique du modèle).

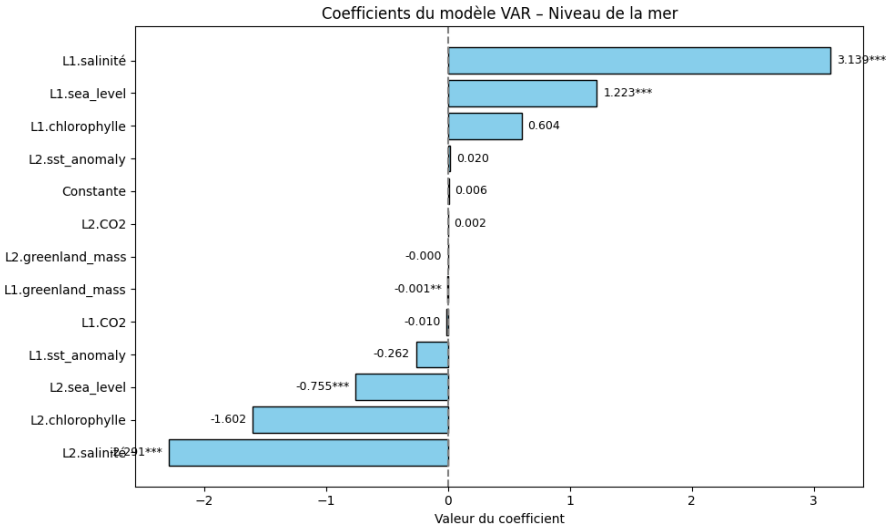

FIGURE 3.10 – Coefficient du modèle VAR(2) - Niveau de la mer différenciée

Les coefficients estimés sur la période 2010–2022 révèlent des comportements économétriques en cohérence avec les processus physiques.

Le modèle VAR(2), estimé sur des séries différenciées, permet d'analyser les variations mensuelles du niveau de la mer en fonction des changements intervenus dans les autres variables du système climatique. L'équation associée aux variations du niveau de la mer met en évidence une dépendance forte et significative à ses propres variations passées (coeff. $L1 = 1.223$, $L2 = -0.755$, $p < 0.001$), ce qui est cohérent avec l'inertie physique des systèmes océaniques documentée dans la littérature (Church & White, 2011). Les variations de la salinité ont un effet significatif de court terme : une variation positive soudaine de la salinité (L1) tend à s'accompagner d'une variation positive du niveau de la mer, suivie d'un effet négatif à L2, ce qui peut refléter un mécanisme transitoire de redistribution océanique. De plus, une diminution rapide de la masse du Groenland (variation négative) est associée à une augmentation immédiate des variations du niveau marin, ce qui soutient l'hypothèse d'un apport d'eau douce contribuant à l'élévation progressive du niveau de la mer (Velicogna et al., 2014). En revanche, les variations du CO_2 atmosphérique et de la température de surface ne montrent pas d'effet direct significatif sur les fluctuations mensuelles du niveau de la mer. Cela peut s'expliquer par le fait que leurs impacts sont de nature indirecte, différée ou non linéaire, et donc moins

détectables dans un modèle linéaire différencié à court terme.Le test de causalité de Granger indique que les anomalies de température de surface de la mer, la perte de masse du Groenland et de l'Antarctique, ainsi que les concentrations de CO_2 Granger-causent le niveau de la mer. Ces résultats sont en accord avec les connaissances établies en climatologie et les résultats obtenus ci-dessus. Notons que le modèle ici présenté ne contient pas la variable de perte de masse de l'Antarctique car le modèle réduit (sans antarctica) présentait de meilleurs résultats ($AIC = -15.32$, $BIC = -13.79$, $HQIC = -14.7$) que le modèle complet le contenant ($AIC = -6.21$, $BIC = -4.14$, $HQIC = -5.37$), et permet une expression un peu plus simplifiée du modèle.

Le modèle affiche un bon ajustement sur les données historiques ($R = 0.82$ pour le niveau de la mer), et les diagnostics de stabilité (racines dans le cercle unité) confirment la validité du système, bien que les tests d'auto-corrélation résiduelle et de normalité indiquent certaines limites, notamment une auto-corrélation pour deux variables (niveau et température de la mer) et une non-normalité de certains résidus. Ces écarts, courants en données climatiques, n'invalident pas le modèle, qui reste pertinent pour l'analyse des dynamiques multivariées. Sur la base de cette spécification, des prévisions à 12 mois ont été générées pour 2023. Les résultats montrent une tendance à la hausse du niveau marin (variant entre $9.06m$ et $13.14m$), une hausse graduelle des températures de surface et une perte de masse glaciaire. Sur la période janvier-juin 2023 (pour laquelle nous disposons des valeurs observées pour sea_level) nous observons un écart absolu moyen de $0,37$ unité par rapport aux valeurs réelles. Cet écart modéré suggère une bonne performance du modèle VAR pour cette variable océanique.

Ces trajectoires sont conformes aux scénarios du GIEC et aux mesures satellitaires (Velicogna et al., 2020). Les graphiques ci-dessous synthétisent l'évolution du niveau de la mer observé, reconstruit et prévu sur l'ensemble de la période.

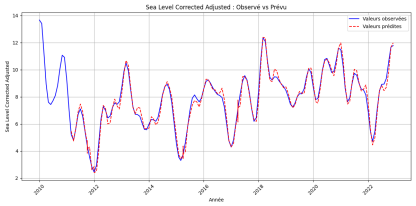

FIGURE 3.11 – Évolution observée et prédite du niveau de la mer

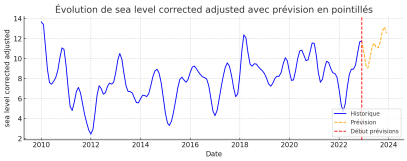

FIGURE 3.12 – Prévisions du niveau de la mer (2023)

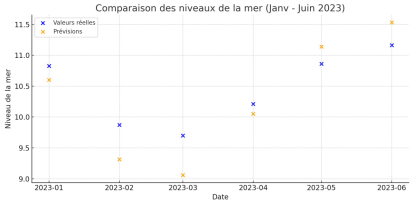

FIGURE 3.13 – Prévisions Vs Vraies valeurs du niveau de la mer (01-06 2023)

Ce modèle VAR, simple mais rigoureusement construit, fournit ainsi un cadre robuste pour la prévision à court terme de variables climatiques interconnectées. Il peut également être utilisé dans des simulations de scénarios prospectifs intégrant des trajectoires d'émission et des hypothèses d'adaptation régionales.

On peut cependant émettre une critique par rapport au sur-apprentissage. Même si l'ordre du modèle a été choisi via les critères AIC et HQIC pour limiter le sur-apprentissage, notre modèle comporte environ $6^2 * 2 = 72$ paramètres pour environ 140 observations ($nb\_parametres \approx dimension\_vecteur^2 * ordre\_modele$).

3.4 Modèle VECM

3.4.1 Spécification du modèle

En nous basant cette fois-ci sur les résultats des tests de stationnarité et du test de cointégration de Johansen, nous allons désormais estimer un modèle vectoriel à correction d'erreur (VECM). Il s'agit d'une extension du modèle VAR adapté aux séries non stationnaires mais cointégrées. Il permet de modéliser à la fois les dynamiques de court terme entre plusieurs séries et les relations d'équilibre de long terme qui les lient. Dans ce modèle, dans un premier temps, seules les variables I(1) sont retenues comme variables explicatives.

Pour estimer un modèle VECM, il faut commencer par choisir 2 paramètres : le nombre de retards (i.e. le nombre de périodes du passées utilisées pour prédire la période future) et le nombre de relations de cointégration entre les variables I(1). Les fonctions python permettant de trouver les paramètres optimaux dépendent elles-même d'un paramètre indiquant si on souhaite ajouter une constante ou une tendance déterministe dans le modèle. Au vu des séries en niveau et des séries différenciées, nous avons retenu 'co' comme choix pour ce paramètre, ce qui correspond à l'ajout d'une constante dans les relations de cointégration mais pas de constante ni de tendance dans les dynamiques de court terme. Une fois ce choix effectué on peut chercher les meilleurs paramètres de retard et de rang de cointégration.

VECM Order Selection (* highlights the minimums)

	AIC	BIC	FPE	HQIC
0	9.250	9.777	1.040e+04	9.464
1	6.855	7.733	950.0	7.212
2	6.014	7.244	410.7	6.514
3	4.991	6.571	148.0	5.633
4	4.415	6.347*	83.79	5.200*
5	4.289	6.571	74.45	5.216
6	4.205*	6.839	69.26*	5.275
7	4.231	7.216	72.23	5.444
8	4.339	7.675	82.09	5.695
9	4.405	8.092	89.91	5.903
10	4.419	8.457	94.01	6.060
11	4.445	8.835	100.3	6.229
12	4.575	9.316	119.5	6.501

FIGURE 3.14 – Choix du meilleur lag ('co')

```
Rang de cointégration selon le lag choisi ('co'):
================================================
Le rang de cointégration pour un lag de 1 est : 2
Le rang de cointégration pour un lag de 2 est : 0
Le rang de cointégration pour un lag de 3 est : 2
Le rang de cointégration pour un lag de 4 (lag optimal) est : 0
Le rang de cointégration pour un lag de 5 est : 1
Le rang de cointégration pour un lag de 6 (lag optimal) est : 1
Le rang de cointégration pour un lag de 7 est : 0
Le rang de cointégration pour un lag de 8 est : 0
```

FIGURE 3.15 – Choix du meilleur rang de cointégration ('co')

Les sorties poussent à choisir un retard de 6 périodes et un rang de cointégration de 1. On peut remarquer qu'on retombe sur les résultats du test de Johansen mené plus haut, ce qui est bon signe.

3.4.2 Interprétation des résultats

En regardant le résumé du modèle disponible en annexe, on observe qu'à part la constante et les coefficients des premiers retard de *sea_level*, tout les autres coefficients sont non significatifs aux seuils habituels. Cela pourrait signifier qu' à court terme, les covariables n'ont pas d'effet sur le niveau de la mer et qu'elles n'influent que via la relation de long terme. Mais étant donné nos connaissances sur la physique de la montée de la mer, ces résultats pourraient plutôt être imputés à une mauvaise spécification du modèle, bien que les paramètres aient été rigoureusement choisis.

La relation de long terme donnée par le modèle est la suivante : $\text{sea\_level}_t + 0.6686 \cdot \text{sea\_temperature}_t + 0.0025 \cdot \text{greenland\_mass}_t - 0.0012 \cdot \text{antarctica\_mass}_t = \mu_t$, où μ_t est une constante (vis-à-vis des variables, non pas du temps). Cette relation semble elle aussi étonnante, car les signes des coefficients de *sea_temperature* et de *antarctica_mass* sont opposés à ce

qu'on pourrait attendre. Mais ces deux coefficients ne sont pas du tout significatifs donc le problème pourrait venir plutôt de la spécification du modèle ou d'autres phénomènes non pris en compte.

Les coefficients de vitesse d'ajustement α montrent comment les variables réagissent à un choc sur la relation de long terme pour revenir à la relation d'équilibre. Encore une fois, seul le coefficient de *sea_level* est significatif. Cela signifie que les autres variables ne s'adaptent pas à un choc et que c'est le niveau de la mer qui porte le poids de l'ajustement vers l'équilibre. Cela semble logique vis-à-vis des variables dans la base de données et des phénomènes physiques sous-jacents. Le coefficient α de *sea_level* est $-0,0514$, ce qui signifie que lorsque le niveau de la mer s'écarte de sa relation d'équilibre à long terme avec les autres variables, il se corrige d'environ $5,14\%$ par période pour revenir à cet équilibre.

Les tests usuels de validation du modèle donnent également des résultats contradictoires. Le test de Ljung-Box de non-corrélation donne des résultats satisfaisants, car les p-valeurs sont supérieures à 5%.On ne rejette donc pas les hypothèses supposant qu'il n'y a pas d'auto-corrélation significative jusqu'au retard h. Mais le test de Jarque-Bera rejette l'hypothèse de normalité des résidus, ce qui suggère que la distribution des erreurs s'écarte de la loi normale. Ce résultat peut remettre en question la validité des inférences statistiques fondées sur le modèle, notamment celles reposant sur l'hypothèse de normalité, comme les tests de significativité des coefficients ou les intervalles de confiance.

Test de Ljung-Box :

	lag_1	lag_2	lag_3	lag_4	lag_5	lag_6	lag_7	lag_8	lag_9	lag_10	lag_11	lag_12
sea_level	0.954	0.802	0.912	0.874	0.798	0.732	0.818	0.886	0.921	0.935	0.962	0.971
sea_temperature	0.861	0.771	0.891	0.217	0.307	0.263	0.172	0.241	0.206	0.262	0.140	0.106
greenland_mass	0.987	0.957	0.993	0.993	0.994	0.732	0.668	0.745	0.823	0.863	0.790	0.628
antarctica_mass	0.643	0.798	0.733	0.864	0.898	0.947	0.880	0.775	0.829	0.876	0.910	0.925

FIGURE 3.16 – Modèle VECM - test de Ljung-Box

```
Test de Jarque-Bera :
==========================================
p-valeur : 2.0136533117828865e-119 < 0.05
On rejete H0, les résidus ne suivent pas une loi normale
```

FIGURE 3.17 – Modèle VECM - test de Jarque-Bera

Cela pourrait expliquer les prédictions faites par le modèle et présentées ci-dessous. Les prédictions proposées pour l'année 2023 semblent en total décalage avec les tendances actuelles et ne nous paraissent pas pertinentes. Cette théorie se vérifie en comparant la prévision faite sur les données de 2020 en n'utilisant que les données antérieures pour l'entraînement. Les prédictions sont en total désaccord avec les valeurs observées sur la même période.
Ces prédictions, ainsi que les résultats des tests de significativité des coefficients et des tests de validation du modèle nous poussent à dire que ce modèle n'est pas adapté pour la prévision du niveau de la mer.
Nous avons comparé les résultats en changeant les paramètres de retard, de rang de cointégration et le paramètre que l'on avait fixé à 'co'. Pour chaque combinaison testée, les résultats obtenus sont similaires à ceux présentés ici, ce qui nous pousse à dire que le problème ne vient pas forcément d'une mauvaise spécification du modèle mais peut-être d'un manque de compréhension de ce modèle de notre part.

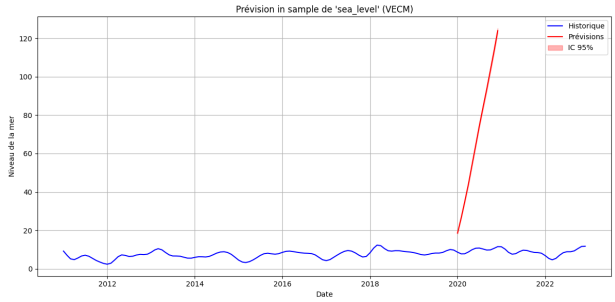

FIGURE 3.18 – Modèle VECM - prévisions in sample

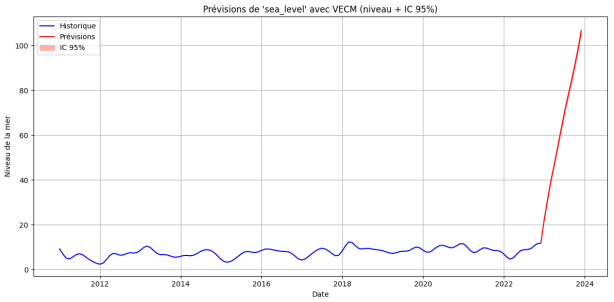

FIGURE 3.19 – Modèle VECM - prévisions out of sample

3.5 Conclusion sur le choix des modèles

3.5.1 Choix pour la modélisation

Au vu des résultats précédents, le modèle VAR(2) nous semble être le modèle le plus adapté pour prédire le futur du niveau des mers. Il prend en compte la dimension temporelle de nos données et obtient de bons scores lors des prédictions. Le seul bémol est la possibilité de sur-apprentissage du au faible ratio entre observations et paramètres, mais ce problème est lié à notre base de données et on peut raisonnablement penser qu'avec une base de plus grande taille ce problème s'estompe.
Le modèle VECM, qui en théorie est adapté à ce genre de situations, s'est avéré faire de très mauvaises prédictions. Il est fort possible que le problème vienne de notre implémentation. En effet nous ne disposons pas encore d'assez de recul sur ce type de modèle que nous venons seulement d'aborder en cours de séries temporelles, ce qui peut expliquer une mauvaise spécification ou une erreur méthodologique dans l'implémentation et l'estimation du modèle.
Par ailleurs, la littérature montre que le modèle VAR performe généralement mieux que le VECM en prévisions car il impose moins de contraintes structurelles sur les relations entre variables. Contrairement au VECM, qui force un retour vers des équilibres de long terme via la cointégration, le VAR se concentre uniquement sur les dynamiques de court terme. Cela le rend plus flexible, notamment lorsque les relations de cointégration sont mal spécifiées ou instables dans le temps. De plus, les erreurs de prévision cumulées dans le VECM, dues aux différences successives, peuvent amplifier les biais. Ainsi, pour des horizons de prévision courts, le VAR s'avère souvent plus robuste et performant.

3.5.2 Pistes d'amélioration

La prochaine piste que nous comptions explorer pour améliorer encore les prédictions aurait été d'inclure des termes de degré 2 dans les modèles de séries temporelles pour tenter de capturer encore plus précisément les interactions en jeu. Mais nous nous serions encore heurtés à des soucis de sur-apprentissage du fait de la taille de notre base de données. L'ajout des relations plus complexes dans le modèle VECM aurait par exemple permis d'obtenir une relation de long terme plus crédible qu'une simple relation linéaire.

Un autre angle d'approche, que nous n'avons pas décidé d'adopter dans ce projet pour des raisons d'interprétabilité, aurait été de nous pencher sur des méthodes non linéaires telles que les algorithmes de forêts aléatoires ou encore les réseaux de neurones. Là encore, notre jeu de données particulièrement restreint aurait entravé cette approche.

Conclusion

Le but de ce projet était d'explorer à l'aide d'une étude statistiques les mécanismes liés à l'élévation du niveau de la mer, avec un focus particulier sur la région de la Méditerranée. La littérature sur le sujet nous a amené à porter notre étude sur les facteurs suivants : température de l'eau, fonte des glaces, salinité, chlorophylle, et concentration en CO_2 dans l'atmosphère. Rappelons ici que les principaux facteurs influents sur la montée du niveau des océans sont la dilatation thermique et la fonte des glaces. Toutefois, ces facteurs proviennent de causes distinctes que nous avons souhaité intégrer dans le jeu de données. Par ailleurs, certains d'entre eux jouent aussi un rôle plus déterminant dans la vulnérabilité côtière face à l'élévation du niveau de la mer, à l'image du taux de chlorophylle. Nous n'avons pas eu le temps dans notre analyse d'évaluer ce niveau de vulnérabilité mais c'est une piste intéressante de travail.

Une première phase de notre travail a constitué à récupérer les données et à créer une base de donnée utilisable et propre.

La phase de modélisation a mis en lumière les limites de certaines approches classiques, telles que la régression multiple, face à la complexité des dynamiques océaniques. Les résultats étant peu significatifs, nous avons cherché à étudier des modèles plus adaptés à la dimension temporelle de nos données.Les modèles VAR se sont révélés plus pertinents, offrant une meilleure capacité de prévision malgré certaines contraintes liées à la taille limitée de notre jeu de données. Le modèle VECM, bien que théoriquement adapté, n'a pas produit de résultats satisfaisants, probablement en raison de difficultés d'implémentation et de la faible stabilité des relations de cointégration dans notre cas.

Au final notre modèle prévoit une élévation du niveau de la mer de 1,9 cm en 2023, soit une valeur significativement plus élevée que les estimations classiques de la littérature, qui tournent autour de 2mm/an. Cette prédiction, bien que cohérente avec notre jeu de données, reflète une dynamique plus marquée. Cette différence s'explique en partie par la forte variabilité des données utilisées. Cela pose la question de la représentativité temporelle et spatiale des données d'entrée : si elles traduisent une réalité régionale spécifique, le résultat n'est pas nécessairement généralisable. À l'inverse, si cette tendance s'inscrit dans un changement structurel, elle pourrait annoncer une sous-estimation systémique par les modèles linéaires traditionnels.

Ce travail confirme que la montée du niveau des mers n'est pas uniquement un enjeu scientifique, mais un véritable défi multidimensionnel, environnemental, social et économique. Il souligne également la nécessité d'une modélisation plus fine, reposant sur des jeux de données plus riches et des méthodes capables de capter la non-linéarité des phénomènes à l'œuvre. Pour aller plus loin, l'exploration de modèles plus complexes ou non linéaires, à condition de disposer de données en quantité suffisante, pourrait constituer une perspective prometteuse.

En somme, ce projet constitue une première étape vers une meilleure compréhension et anticipation de la montée des eaux en Méditerranée. Il illustre à la fois les potentialités et les défis d'une approche statistique appliquée à des questions climatiques urgentes et concrètes.

Bibliographie

[1] Stephen Kaczkowski and. Mathematical models for global mean sea level rise. *The College Mathematics Journal*, 48(3) :162–169, 2017.

[2] Jean-Joseph Blanc and Hugues Faure. La montée récente du niveau de la mer. exemples de marseille, gênes et venise (méditerranée). *Géologie Méditerranéenne*, 17(2) :109–122, 1990.

[3] Anny Cazenave and Gonéri Le Cozannet. Sea level rise and its coastal impacts. *Earth's Future*, 2(2) :15–34, 2014.

[4] Copernicus Marine Service. Copernicus marine environment monitoring service (cmems), 2024.

[5] Copernicus Marine Service. $\mathrm{Omi}_{h}ealth_{c}hl_{m}edsea_{o}ceancolour_{t}rend$: $Mediterraneanseachlorophyllconcentrationtrendfromoceancolourobservations$, 2024.

[6] Copernicus Marine Service. Sea level – ocean climate portal. `https://marine.copernicus.eu/fr/ocean-climate-portal/sea-level`, 2024.

[7] Gilles Delaygue, Jean Jouzel, Jean-François Minster, Jean-Louis Dufresne, Olivier Boucher, and Marie-Antoinette Mélière. La fonte des glaces et l'élévation du niveau marin. *Planet Terre*, 05 2001.

[8] European Space Agency. Esa earth observation data, 2024.

[9] Thomas Frederikse, Felix Landerer, Lambert Caron, Surendra Adhikari, David Parkes, Vincent W Humphrey, Sönke Dangendorf, Peter Hogarth, Laure Zanna, Lijing Cheng, et al. The causes of sea-level rise since 1900. *Nature*, 584(7821) :393–397, 2020.

[10] IPCC. *Special Report on the Ocean and Cryosphere in a Changing Climate*. Intergovernmental Panel on Climate Change, 2019. Synthesis Report on the Ocean and Cryosphere in a Changing Climate.

[11] IPCC. *Climate Change 2022 : Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK and New York, NY, USA, 2022.

[12] Alix Lombard. *Les variations actuelles du niveau de la mer : Observations et causes.* Theses, Université Paul Sabatier - Toulouse III, November 2005.

[13] Alix Lombard. Les variations actuelles du niveau de la mer : observations et causes climatiques [prix prud'homme 2006]. *La Météorologie*, 2007(59) :13–21, 2007.

[14] Marta Marcos, Michael N. Tsimplis, and Andrew G. P. Shaw. 21st century mediterranean sea level rise : Steric and atmospheric pressure contributions from a regional model. *Global and Planetary Change*, 62(3-4) :189–209, 2008.

[15] NASA. Nasa earth observing system data and information system (eosdis), 2024.

[16] NOAA. Noaa national centers for environmental information (ncei), 2024.

[17] NOAA Global Monitoring Laboratory. Noaa esrl gml : Carbon cycle greenhouse gases - flask data, 2024.

[18] Copernicus Marine Service. Mediterranean sea mean sea level time series and trend from observations reprocessing. 2024.

[19] Saumya Sinha, John Fasullo, R Steven Nerem, and Claire Monteleoni. Sea level projections with machine learning using altimetry and climate model ensembles. *arXiv preprint arXiv :2308.02460*, 2023.

[20] Michael N Tsimplis, Marta Marcos, and Samuel Somot. 21st century mediterranean sea level rise : Steric and atmospheric pressure contributions from a regional model. *Global and Planetary Change*, 63(2-3) :105–111, 2008.

Chapitre 4

Annexe

4.1 Lien vers le dépot Github

`https://github.com/LeoDony7/Stat-App-Montee-des-oceans`

4.2 Extrait de la base de données (Année 2015)

year_month	sea_level	sea_temperature	greenland_mass	antarctica_mass	chlorophylle	CO2	sea_salinity
2015-01-01	4.557668	0.450087	-3493.910000	-1772.640000	0.019114	0.685000	0.008140
2015-02-01	3.555065	0.458027	-3502.230000	-1859.940000	0.016359	2.532500	-0.000638
2015-03-01	3.304771	0.462972	-3456.580000	-1826.780000	0.027040	-0.350000	-0.005128
2015-04-01	3.741735	0.469873	-3449.320000	-1810.790000	0.023347	2.485000	-0.014665
2015-05-01	4.643012	0.472627	-3508.706667	-1775.663333	0.004663	1.767500	-0.021187
2015-06-01	5.793702	0.473899	-3568.093333	-1740.536667	0.002634	1.980833	0.009270
2015-07-01	6.998950	0.490419	-3627.480000	-1705.410000	0.000288	5.314000	0.010155
2015-08-01	7.893311	0.502378	-3762.110000	-1859.760000	-0.000273	4.095000	-0.073452
2015-09-01	8.141421	0.498321	-3801.600000	-1820.900000	0.002700	1.567500	-0.071438
2015-10-01	7.864274	0.493246	-3781.610000	-1827.636667	0.005594	1.793333	-0.060363
2015-11-01	7.629857	0.494104	-3761.620000	-1834.373333	-0.001258	5.035833	0.001826
2015-12-01	7.911013	0.492411	-3741.630000	-1841.110000	-0.001279	1.290000	-0.023072

TABLE 4.1 – Extrait du jeu de données (Année 2015)

4.3 Modèle physique de la hauteur de la mer - Calculs

Le modèle repose sur les hypothèses suivantes :
- La température et la salinité des eaux de surface des océans sont uniformes.
- Les eaux de surface sont supposées immobiles (absence de courants).
- La profondeur minimale des océans est supérieure à une valeur constante H.
- Le réchauffement ne concerne qu'une couche d'eau de profondeur H qui passe à une température supérieure T^*.
- La Terre est modélisée comme une sphère parfaite de rayon R, dont une fraction f de la surface est couverte par les océans.
- La variation du niveau marin est modélisée par une élévation moyenne h.

Le volume initial V des eaux de surface est donné par :

$$V = 4\pi R^2 H f. \tag{4.1}$$

Après réchauffement, la densité moyenne diminue de ρ à ρ^*, et le nouveau volume devient :

$$V^* \approx 4\pi R^2 (H + h) f. \tag{4.2}$$

En conservant la masse ($\rho V = \rho^* V^*$), on obtient une expression de l'élévation moyenne du niveau marin due à la dilatation thermique :

$$h = H \left(\frac{\rho}{\rho^*} - 1 \right). \tag{4.3}$$

Cette élévation h représente une moyenne globale d'élévation sensible aux variations, même faibles, de densité.

Soit m_I la masse de glace fondue. Le volume d'eau liquide obtenu est :

$$V = \frac{m_I}{\rho_w}, \tag{4.4}$$

où ρ_w est la densité de l'eau (environ $1000\,\mathrm{kg/m^3}$).

Le nouveau volume total des océans devient :

$$V^* = V + \Delta V = 4\pi R^2 (H + h) f. \tag{4.5}$$

En isolant l'élévation h due à la fonte des glaces, on obtient :

$$h = \frac{m_I}{4\pi R^2 \rho_w f}. \tag{4.6}$$

4.4 Spécification du modèle VAR

1. Choix du VAR(2)

TABLE 4.2 – Critères d'information pour les modèles VAR(1) et VAR(2)

Modèle	**AIC**	**BIC**	**HQIC**
VAR(1)	-13,76	-12,937	-13,43
VAR(2)	**-15,32**	**-13,79**	**-14,7**

TABLE 4.3 – Erreur quadratique moyenne de prévision (RMSE)

Modèle	**RMSE**
VAR(1)	**32,57**
VAR(2)	34,10

TABLE 4.4 – Comparaison des coefficients de détermination (R^2 et R^2 ajusté) pour VAR(1) et VAR(2)

Variable	**VAR(1)**		**VAR(2)**	
	R^2	R^2 ajusté	R^2	R^2 ajusté
Sea level corrected adjusted	0.5296	0.5105	**0.8267**	**0.8119**
SST anomaly filtered	0.6171	0.6016	**0.6727**	**0.6448**
Greenland mass	0.2898	0.2610	**0.4184**	**0.3689**
CO_2 seasonal	0.2355	0.2045	**0.3606**	**0.3062**
Chlorophylle seasonal	0.6104	0.5946	**0.7198**	**0.6959**
Salinité seasonal	0.7242	0.7131	**0.8056**	**0.7891**

2. Notation

Soit $y_t^{(i)}$ la variable i-ème de notre modèle, où

$$i \in \{\text{sea\_level, sea\_temperature, greenland\_mass,}$$

$$\text{CO}_2\text{, chlorophylle, sea\_salinity}\}$$

Soit $\Delta y_t^{(i)}$ la première différence de la variable i-ème, c'est-à-dire $\Delta y_t^{(i)} = y_t^{(i)} - y_{t-1}^{(i)}$.

Soit $\Delta^{12} y_t^{(i)}$ la différence saisonnière d'ordre 12, c'est-à-dire $\Delta^{12} y_t^{(i)} = y_t^{(i)} - y_{t-12}^{(i)}$.

Les termes d'erreur $\varepsilon_t^{(i)}$ sont définis comme suit : $\varepsilon_t^{(i)}$ est le terme d'erreur de l'équation i-ème, à l'instant t.

Ainsi, le modèle VAR s'écrit :

$$\Delta \mathbf{y}_t = \mathbf{c} + A_1 \Delta \mathbf{y}_{t-1} + A_2 \Delta \mathbf{y}_{t-2} + \varepsilon_t$$

(1)

Avec :

$$\Delta \mathbf{y}_t = \begin{pmatrix} \Delta y_t^{(1)} \\ \Delta y_t^{(2)} \\ \Delta y_t^{(3)} \\ \Delta^{12} y_t^{(4)} \\ \Delta^{12} y_t^{(5)} \\ \Delta^{12} y_t^{(6)} \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \end{pmatrix} = \begin{bmatrix} 0.006519 \\ 0.001362 \\ -16.227465 \\ 2.126733 \\ 0.000459 \\ 0.001315 \end{bmatrix}$$

$$A_1 = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{pmatrix} = \begin{bmatrix} 1.223143 & -0.262089 & -0.001065 & -0.009692 & 0.604230 & 3. \\ 0.005429 & 1.005700 & 0.000017 & 0.000093 & 0.019087 & -0 \\ 5.098379 & -102.326656 & 0.611352 & -0.045950 & -172.580019 & 97 \\ 0.326269 & 3.241663 & 0.006728 & 0.093711 & 51.50922 & 4 \\ 0.000085 & 0.150236 & -0.000009 & -0.000215 & 0.937545 & 0. \\ -0.005591 & 0.058649 & 0.000015 & 0.001237 & 0.258234 & 1. \end{bmatrix}$$

$$A_2 = \begin{pmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} & b_{16} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} & b_{26} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} & b_{36} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} & b_{46} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} & b_{56} \\ b_{61} & b_{62} & b_{63} & b_{64} & b_{65} & b_{66} \end{pmatrix} = \begin{bmatrix} -0.755013 & 0.020084 & -0.000239 & 0.001927 & -1.602122 & -2. \\ -0.003507 & -0.308868 & -0.000019 & -0.000157 & -0.079093 & 0.0 \\ -13.886163 & 3329.142895 & -0.306313 & 0.991767 & 544.709476 & -32 \\ -0.605864 & -13.924813 & -0.008353 & -0.031463 & -13.712882 & -35 \\ -0.001604 & -0.166636 & 0.000036 & 0.000077 & -0.28095 & -0 \\ 0.003462 & -0.242011 & 0.000087 & -0.001117 & 0.285296 & -0. \end{bmatrix}$$

$$\varepsilon_t = \begin{pmatrix} \varepsilon_{\text{sea\_level\_diff},t} \\ \varepsilon_{\text{sea\_temperature\_diff},t} \\ \varepsilon_{\text{greenland\_mass\_diff},t} \\ \varepsilon_{\text{CO}_2\text{\_seasonal\_diff},t} \\ \varepsilon_{\text{chlorophylle\_seasonal\_diff},t} \\ \varepsilon_{\text{sea\_salinity\_seasonal\_diff},t} \end{pmatrix}$$

3. Passage en niveau

En effectuant les développements qu'il faut, nous aboutissons à :

$$(1) \quad \Longleftrightarrow \quad \begin{pmatrix} \Delta y_t^{(1)} \\ \Delta y_t^{(2)} \\ \Delta y_t^{(3)} \\ \Delta^{12} y_t^{(4)} \\ \Delta^{12} y_t^{(5)} \\ \Delta^{12} y_t^{(6)} \end{pmatrix} = \mathbf{c} + A_1 \begin{pmatrix} \Delta y_{t-1}^{(1)} \\ \Delta y_{t-1}^{(2)} \\ \Delta y_{t-1}^{(3)} \\ \Delta^{12} y_{t-1}^{(4)} \\ \Delta^{12} y_{t-1}^{(5)} \\ \Delta^{12} y_{t-1}^{(6)} \end{pmatrix} + A_2 \begin{pmatrix} \Delta y_{t-2}^{(1)} \\ \Delta y_{t-2}^{(2)} \\ \Delta y_{t-2}^{(3)} \\ \Delta^{12} y_{t-2}^{(4)} \\ \Delta^{12} y_{t-2}^{(5)} \\ \Delta^{12} y_{t-2}^{(6)} \end{pmatrix} + \varepsilon_t$$

$$\iff \quad \mathbf{y}_t = \begin{pmatrix} y_t^{(1)} \\ y_t^{(2)} \\ y_t^{(3)} \\ y_t^{(4)} \\ y_t^{(5)} \\ y_t^{(6)} \end{pmatrix} = \mathbf{c} + A_1 \begin{pmatrix} y_{t-1}^{(1)} - y_{t-2}^{(1)} \\ y_{t-1}^{(2)} - y_{t-2}^{(2)} \\ y_{t-1}^{(3)} - y_{t-2}^{(3)} \\ y_{t-1}^{(4)} - y_{t-13}^{(4)} \\ y_{t-1}^{(5)} - y_{t-13}^{(5)} \\ y_{t-1}^{(6)} - y_{t-13}^{(6)} \end{pmatrix} + A_2 \begin{pmatrix} y_{t-2}^{(1)} - y_{t-3}^{(1)} \\ y_{t-2}^{(2)} - y_{t-3}^{(2)} \\ y_{t-2}^{(3)} - y_{t-3}^{(3)} \\ y_{t-2}^{(4)} - y_{t-14}^{(4)} \\ y_{t-2}^{(5)} - y_{t-14}^{(5)} \\ y_{t-2}^{(6)} - y_{t-14}^{(6)} \end{pmatrix} + \varepsilon_t$$

Nous obtenons ainsi la valeur prédite :

$$\widehat{Y_t} \ = \mathbf{c} + A_1 \begin{pmatrix} y_{t-1}^{(1)} - y_{t-2}^{(1)} \\ y_{t-1}^{(2)} - y_{t-2}^{(2)} \\ y_{t-1}^{(3)} - y_{t-2}^{(3)} \\ y_{t-1}^{(4)} - y_{t-13}^{(4)} \\ y_{t-1}^{(5)} - y_{t-13}^{(5)} \\ y_{t-1}^{(6)} - y_{t-13}^{(6)} \end{pmatrix} + A_2 \begin{pmatrix} y_{t-2}^{(1)} - y_{t-3}^{(1)} \\ y_{t-2}^{(2)} - y_{t-3}^{(2)} \\ y_{t-2}^{(3)} - y_{t-3}^{(3)} \\ y_{t-2}^{(4)} - y_{t-14}^{(4)} \\ y_{t-2}^{(5)} - y_{t-14}^{(5)} \\ y_{t-2}^{(6)} - y_{t-14}^{(6)} \end{pmatrix}$$

Pour prévoir Y_{t+1}, on répète l'opération en utilisant la valeur prédite $\widehat{Y_t}$, etc.

4. Résultats des différents tests

TABLE 4.5 – Résultats du test de Ljung-Box (lag = 2)

Variable	**Statistique Ljung-Box**	**p-value**
sea level corrected adjusted_diff	17.681	0.00015
sst_anomaly_filtered_diff	7.809	0.02016
greenland_mass_diff	0.115	**0.94423**
CO2_seasonal_diff	0.405	**0.81683**
chlorophylle_seasonal_diff	1.855	**0.39549**
salinité_seasonal_diff	0.483	**0.78563**

TABLE 4.6 – Résultats du test de Jarque-Bera (normalité des résidus)

Variable	Statistique JB	p-value
sea level corrected adjusted_diff	0.713	**0.700**
sst_anomaly_filtered_diff	7.933	0.019
greenland_mass_diff	35.540	0.000
CO2_seasonal_diff	6270.465	0.000
chlorophylle_seasonal_diff	29.308	0.000
salinité_seasonal_diff	42.693	0.000

TABLE 4.7 – Valeurs propres du modèle VAR(1)- algorithme de la matrice compagnon

Valeur propre	Multiplicité
0.2639	2
0.4989	2
0.5563	2
0.6633	2
0.8101	2
0.8489	2

4.5 Modèle VECM - Coefficients estimés

Det. terms outside the coint. relation & lagged endog. parameters for equation sea_level

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.0374	0.017	2.226	0.026	0.004	0.070
L1.sea_level	2.5590	0.083	30.846	0.000	2.396	2.722
L1.sea_temperature	-1.0748	1.166	-0.922	0.357	-3.360	1.211
L1.greenland_mass	4.133e-05	0.000	0.178	0.858	-0.000	0.000
L1.antarctica_mass	-2.119e-05	0.000	-0.150	0.881	-0.000	0.000
L2.sea_level	-3.4023	0.239	-14.208	0.000	-3.872	-2.933
L2.sea_temperature	0.2264	1.873	0.121	0.904	-3.444	3.896
L2.greenland_mass	0.0004	0.000	1.414	0.157	-0.000	0.001
L2.antarctica_mass	-9.957e-05	0.000	-0.722	0.470	-0.000	0.000
L3.sea_level	2.7192	0.369	7.374	0.000	1.996	3.442
L3.sea_temperature	0.4376	2.013	0.217	0.828	-3.509	4.384
L3.greenland_mass	-1.043e-05	0.000	-0.039	0.969	-0.001	0.001
L3.antarctica_mass	-0.0001	0.000	-1.066	0.287	-0.000	0.000
L4.sea_level	-1.1241	0.371	-3.032	0.002	-1.851	-0.398
L4.sea_temperature	-0.1101	2.036	-0.054	0.957	-4.100	3.880
L4.greenland_mass	0.0002	0.000	0.939	0.348	-0.000	0.001
L4.antarctica_mass	6.841e-05	0.000	0.472	0.637	-0.000	0.000
L5.sea_level	0.0939	0.235	0.400	0.689	-0.366	0.554
L5.sea_temperature	-1.7632	1.877	-0.939	0.347	-5.442	1.915
L5.greenland_mass	0.0002	0.000	0.712	0.476	-0.000	0.001
L5.antarctica_mass	-4.428e-05	0.000	-0.301	0.763	-0.000	0.000
L6.sea_level	0.1281	0.085	1.510	0.131	-0.038	0.294
L6.sea_temperature	0.7734	1.167	0.663	0.508	-1.514	3.061
L6.greenland_mass	1.378e-05	0.000	0.062	0.951	-0.000	0.000
L6.antarctica_mass	-1.235e-05	0.000	-0.084	0.933	-0.000	0.000

FIGURE 4.1 – Modèle VECM - coefficients de court terme pour *sea_level*

Loading coefficients (alpha) for equation sea_level

	coef	std err	z	P>|z|	[0.025	0.975]
ec1	-0.0514	0.011	-4.749	0.000	-0.073	-0.030

Loading coefficients (alpha) for equation sea_temperature

	coef	std err	z	P>|z|	[0.025	0.975]
ec1	-0.0011	0.001	-1.461	0.144	-0.003	0.000

Loading coefficients (alpha) for equation greenland_mass

	coef	std err	z	P>|z|	[0.025	0.975]
ec1	-4.8539	3.995	-1.215	0.224	-12.685	2.977

Loading coefficients (alpha) for equation antarctica_mass

	coef	std err	z	P>|z|	[0.025	0.975]
ec1	-3.7163	6.073	-0.612	0.541	-15.618	8.186

Cointegration relations for loading-coefficients-column 1

	coef	std err	z	P>|z|	[0.025	0.975]
beta.1	1.0000	0	0	0.000	1.000	1.000
beta.2	0.6686	3.382	0.198	0.843	-5.960	7.297
beta.3	0.0025	0.001	2.414	0.016	0.000	0.005
beta.4	-0.0012	0.001	-0.829	0.407	-0.004	0.002

FIGURE 4.2 – Modèle VECM - coefficients de long terme