

Faculté des sciences

Interpolation géospatiale des
observations de températures entre
500 et 1800AD en tenant compte de
simulations climatiques

Approche par processus gaussiens

Auteur : Corentin Visée

Promoteur : Pr. Michel Crucifix

Lecteurs : Pr. Hugues Goosse et Pr. Bas van Wesemael

Année académique 2019-2020

Master [120] en sciences géographiques, orientation climatologie, à finalité approfondie

Remerciements

Je tiens à remercier particulièrement M. Crucifix pour son encadrement, ses conseils et sa patience au cours de ces deux années de travail.

J'adresse également mes remerciements à M. Klein pour toutes les réponses à mes questions ainsi que pour m'avoir fourni le modèle sans lequel ce travail n'aurait pas été possible.

Résumé

La qualité des reconstructions temporelles et spatiales du climat au cours des 2000 dernières années est fortement dépendante de la répartition et de la disponibilité des données climatiques. L'absence d'observations directes des températures au-delà des 150 dernières années nécessite l'utilisation d'observations indirectes, comme les sédiments marins ou lacustres. Ces mesures indirectes, contrairement aux données instrumentales, peuvent fournir des séries temporelles sur une période plus longue. Ces séries sont souvent éparses et non continues. Une régression faisant appel aux processus gaussiens peut toutefois fournir des séries complètes pour l'intervalle 500-1800 *AD*.

Ces données indirectes sont également ponctuellement réparties sur le globe et le passage à un champ continu peut se réaliser en utilisant différentes méthodes (le krigeage ou l'assimilation de données en font partie). La contrainte d'un modèle par des observations est une opération complexe, notamment dans un cadre paléoclimatique. Il est également complexe de fournir un cadre physique exhaustif à une interpolation. Nous proposons dans ce travail une troisième approche originale à mi-chemin entre l'interpolation et l'assimilation de données. Le lien entre les points de grille n'est plus paramétré comme une fonction de la distance. Ce lien est fourni par la corrélation entre les séries temporelles du modèle de complexité intermédiaire *LOVECLIM*. Cette approche est réalisée grâce aux équations prédictives des processus gaussiens. Cette méthode originale permet à un nombre réduit d'observations de fournir une information sur une étendue variable de cellules. Nous développons ici les éléments principaux nécessaires à son application dans un contexte lié au paléoclimat.

Abstract

The quality of temporal and spatial reconstructions of the climate over the past 2000 years is highly dependent on the distribution and availability of climate data. The lack of direct observations of temperatures beyond the last 150 years requires the use of indirect observations, such as marine or lake sediments. These indirect measurements, unlike instrumental data, can provide time series over a longer period. These series are often scattered and non-continuous. A regression using Gaussian processes, however, can provide complete series for the interval 500-1800 *AD*.

These indirect data present a patchy distribution over the globe and the transition to a continuous field can be achieved using various methods (kriging or data assimilation are part of this). The constraint of a model by observations is a complex operation, especially in a paleoclimatic framework. It is also complex to provide a comprehensive physical framework for an interpolation. We propose in this work a third original approach halfway between interpolation and data assimilation. The link between grid points is no longer parameterized as a function of distance. This link is provided by the correlation between the time series of the intermediate complexity model *LOVECLIM*. This approach is carried out using the predictive equations of Gaussian processes. This original method allows a reduced number of observations to provide information on a variable set of location. We develop here the main elements necessary for its application in a context linked to the paleoclimate.

Table des matières

1	Introduction	2
1.1	Réchauffement climatique et paléoclimat	2
1.2	Archives et <i>proxies</i>	3
1.2.1	Types d'archives et <i>proxies</i> associés	3
1.2.2	Reconstruction d'une série temporelle	5
1.2.3	Reconstruction d'un champ spatial	5
2	Matériel et méthode	10
2.1	Matériel	10
2.1.1	Modèle <i>LOVECLIM</i>	10
2.1.2	Archives	11
2.2	Méthode	12
2.2.1	Processus gaussiens et séries temporelles	12
2.2.2	Processus gaussiens et interpolation spatiale	16
2.2.3	Analyse en composantes principales	17
3	Résultats	19
3.1	Séries temporelles	19
3.1.1	Remplissage	19
3.2	Champ spatial	22
3.2.1	Structures de corrélation : <i>LOVECLIM</i>	22
3.2.2	Composantes principales	24
3.2.3	Reconstruction	25
3.2.4	Modes propres de la reconstruction	28
4	Discussion	30
4.1	Séries temporelles	30
4.1.1	Influence du choix d'un modèle exponentiel	30
4.1.2	Influence du bruit sur le modèle exponentiel	31
4.1.3	Influence du type de <i>proxy</i>	32
4.2	Champ spatial	33
4.2.1	Influence de la période considérée pour <i>LOVECLIM</i>	33
4.2.2	Influence du remplissage sur la reconstruction spatiale	34
5	Conclusion	36

Chapitre 1

Introduction

1.1 Réchauffement climatique et paléoclimat

Les changements des concentrations atmosphériques en gaz à effet de serre sont particulièrement rapides depuis le milieu du 18^e siècle. La concentration de dioxyde de carbone a augmenté de 130ppm en moins de deux siècles (280ppm à 410ppm). Les valeurs observées de nos jours sont supérieures à ce qui est observé au cours des 800 000 dernières années (figure 1.1). Cette augmentation est presque exclusivement causée par l'utilisation de combustibles fossiles et par les activités de déboisement massif [IPCC, 2014].

Cette augmentation des concentrations accentue le piégeage radiatif par effet de serre, ce qui augmente les températures globales [IPCC, 2014], entraînant de nombreuses conséquences sur la biodiversité [Malcolm et al., 2006], les ressources en eau [Şen, 2009] et les zones de répartition des maladies à vecteurs [Cook, 1992].

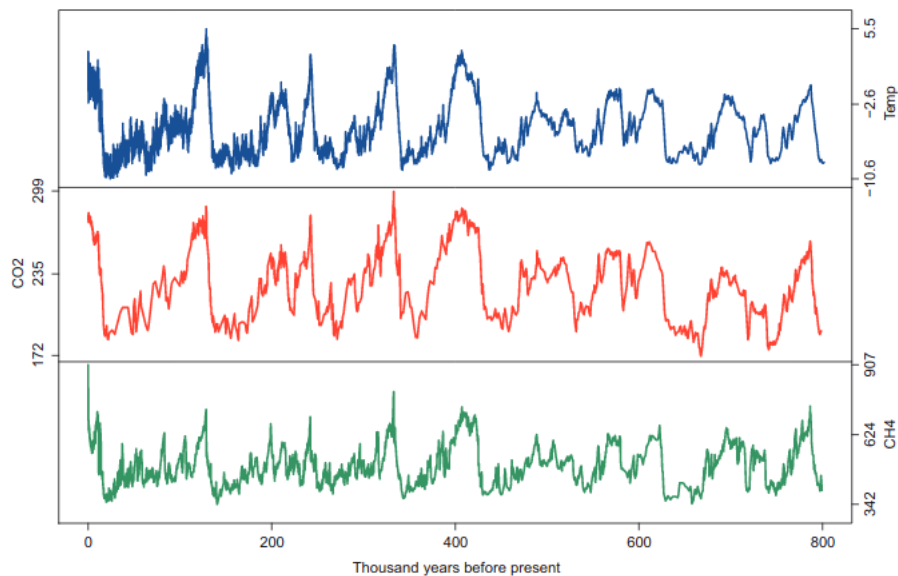


FIGURE 1.1 – Concentrations de différents gaz déterminées par un carottage lors du programme EPICA [Nieto-Barajas, 2018]

Bien qu'il soit admis que ce réchauffement global soit causé par l'activité humaine, l'empreinte anthropique et les fluctuations naturelles du climat ne sont pas toujours faciles à distinguer à l'échelle régionale. Les fluctuations naturelles ont des causes variées [Wratt and Mulan, 2007], telles que la variation des éléments extérieurs (irradiance solaire, éruptions volcaniques,...) ou la dynamique interne du système climatique (oscillation *El-Nino*). Ces fluctuations peuvent s'observer à différentes échelles : annuelles, décennales et au-delà. Il est donc particulièrement important de disposer de séries temporelles de longue durée afin de caractériser ces variations à différentes échelles. Ces séries doivent également être réparties sur le globe.

Les mesures directes sont toutefois limitées aux 150 dernières années. Il n'est donc pas possible de détecter des changements dans la fréquence d'événements extrêmes ou dans la variabilité des oscillations naturelles du climat grâce aux données instrumentales [National Centers for Environmental Information, 2020]. Il est possible de contourner en partie cette limitation en utilisant des archives naturelles du climat (sédiments océaniques ou lacustres, cernes d'arbres,...). Ces archives contiennent des informations sur les conditions climatiques présentes lors de leur formation. Ces informations peuvent être extraites et analysées (figure 1.2).



FIGURE 1.2 – Archives et *proxies* climatiques [USGS, 2020]

L'utilisation de ce type de données mène toutefois à d'autres défis. Ces archives ne sont pas distribuées de façon homogène autour du globe. De plus, la résolution temporelle, la précision et la fiabilité des informations climatiques extraites varient en fonction des types d'archives et même au sein d'un type d'archives. La reconstruction d'un champ spatio-temporel est encore un problème scientifique ouvert pour lequel il n'existe pas de solution unique. Dans ce mémoire, nous présentons une reconstruction pour les 2000 dernières années. La méthode utilisée dans ce mémoire est originale. Pour le montrer, nous allons passer en revue les méthodes plus couramment employées.

1.2 Archives et *proxies*

1.2.1 Types d'archives et *proxies* associés

De nombreux types d'archives climatiques peuvent fournir des informations sur le paléoclimat et seront utilisées dans le cadre de ce mémoire.

Carottages glaciaires

Les calottes polaires et les glaciers continentaux se forment par l'accumulation de différentes couches de neige au cours du temps [Fischer et al., 2004]. Ces archives fournissent des informations sur les conditions atmosphériques passées (notamment grâce aux bulles piégées de CO_2 et de CH_4), sur

l'intensité du vent (grâce aux poussières) ou encore sur les précipitations liées aux températures de surface (grâce aux isotopes de l'oxygène). Ce type de données peut remonter jusqu'aux 800 000 dernières années avec une résolution sous-saisonnière jusque décennale [USGS, 2020].

Les sédiments

Ces archives sont des dépôts sédimentaires présents en couches dans les zones lacustres, océaniques ou encore marécageuses [USGS, 2020].

Pour les sédiments lacustres, l'analyse porte sur les pollens ou les diatomées¹ afin de déterminer des changements paléoenvironnementaux ou climatiques. Les changements de ces éléments peuvent permettre d'extraire des informations sur la température et l'humidité lors du dépôt [Fischer et al., 2004]. Les sédiments présents au fond des lacs possèdent généralement une résolution saisonnière et permettent une étude jusqu'à 150 000 ans [Fischer et al., 2004].

Pour les sédiments marins, les analyses portent principalement sur les isotopes et les microfossiles présents qui permettent de déterminer les températures de surface et des concentrations atmosphériques en CO_2 [Fischer et al., 2004]. La résolution temporelle est assez grossière puisqu'elle varie entre 100 et 1000 ans. La datation au carbone permet de remonter jusque 50 000 ans et il est même possible de remonter jusqu'au million d'années par des méthodes plus complexes (autres isotopes radioactifs ou paléomagnétisme) [Fischer et al., 2004].

Les cernes d'arbres

La croissance des arbres est marquée chaque année par la création d'une nouvelle couche à leur tronc appelée cerne de croissance [USGS, 2020]. Ces cernes, lorsqu'ils sont analysés, permettent de déterminer les températures et les précipitations auxquelles les arbres ont été soumis [Fischer et al., 2004]. En effet, les changements d'épaisseur des cernes indiquent des changements saisonniers et annuels dans la disponibilité en eau [USGS, 2020]. Il est possible de disposer d'une information annuelle sur les 14 000 dernières années avec ce type d'archives [Fischer et al., 2004].

Les coraux

Les coraux sont des organismes marins vivant dans les eaux chaudes peu profondes et possédant un squelette calcaire [USGS, 2020]. La croissance des coraux se marque par une superposition de couches claires en été et sombres en hiver [USGS, 2020]. En analysant ces couches, il est possible d'obtenir une information mensuelle, annuelle et multiannuelle sur la température et la salinité du lieu de croissance [Fischer et al., 2004]. Il est possible d'utiliser ces archives pour les 130 000 dernières années [Fischer et al., 2004].

Les spéléothèmes

Ces archives sont des dépôts créés lors du passage de l'eau dans des grottes. Les changements environnementaux dans les conditions de surface affectent les taux de croissance et les compositions de ces formations géologiques. Les proxies climatiques associés à ces archives sont nombreux et variés. Il peut en effet s'agir de composition isotopique, de contenu en matériaux organiques ou encore d'oligo-éléments [Jones and Mann, 2004]. Bien que la résolution temporelle semble être annuelle, la datation exacte des différentes couches pose encore question aujourd'hui au sein de la communauté scientifique. L'interprétation précise du *proxy* climatique choisi est également complexe car les influences extérieures sont nombreuses sur une variable climatique, notamment la température [Jones and Mann, 2004].

1. Végétal unicellulaire aquatique [Larousse, 2020]

Les documents historiques

Ces sources de données sur le climat sont associées à l'enregistrement des événements climatiques extrêmes tels que les sécheresses, famines ou même parfois les dates de floraison des plantes [Jones and Mann, 2004]. Ces données sont toutefois peu nombreuses et il n'est pas rare que les auteurs exagèrent les conditions extrêmes lors de leur rédaction. Des biais importants sont également causés par la subjectivité de l'observateur et l'interprétation de son vocabulaire.

1.2.2 Reconstruction d'une série temporelle

Nous avons vu que les archives naturelles du climat enregistrent, grâce à des phénomènes biologiques, chimiques ou physiques, des phénomènes liés au climat [Jones and Mann, 2004]. Cet enregistrement d'informations peut être analysé grâce à différentes méthodes, comme par exemple les rapports isotopiques. Ces rapports peuvent ensuite être interprétés comme des indicateurs des conditions de formations (figure 1.3) [Jones and Mann, 2004].

L'utilisation d'indicateurs indirects nécessite une validation et une calibration grâce à des mesures directes de la température [USGS, 2020]. Lorsque ces données instrumentales sont disponibles à proximité de la localisation des archives, il est possible de déterminer un lien entre les données instrumentales et le *proxy* par régression linéaire [Thejll and Schmith, 2005]. Lorsque ces données instrumentales ne sont pas disponibles, il est nécessaire de faire des hypothèses sur le filtrage du signal climatique par les archives. Les modèles sont donc plus complexes et sont sujets à débat [Jones and Mann, 2004].

Deux aspects de la reconstruction sont donc distinguables dans ce travail. Le premier consiste à reconstruire le champ spatial sur base d'observations et le second consiste à reconstruire une évolution temporelle continue sur base d'observations discrètes. Ces deux aspects doivent ensuite être combinés.

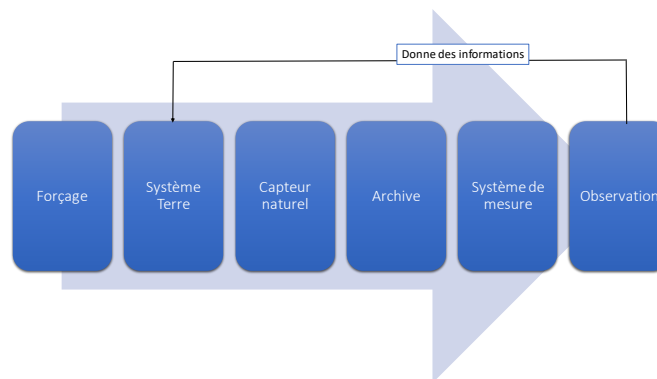


FIGURE 1.3 – Formation d'un *proxy* climatique

1.2.3 Reconstruction d'un champ spatial

Deux grandes approches existent afin de reconstruire des champs spatiaux de températures. La première est l'interpolation d'observations avec des modèles géospatiaux tels que le *krigeage*. La seconde consiste à contraindre des simulations climatiques grâce à des observations par des méthodes d'assimilation de données [Widmann et al., 2010].

Le but de ce mémoire est de proposer une troisième option permettant de combiner les avantages des méthodes d'interpolation et des méthodes d'assimilation de données. Les prochains paragraphes ont pour objectif de présenter ces deux méthodes avant d'expliquer l'intérêt d'une troisième.

Interpolation spatiale

Les méthodes d'interpolation sont généralement utilisées pour réaliser un champ continu d'une variable en appliquant une pondération en fonction de la distance séparant les données ponctuelles [Attorre et al., 2007]. Il existe différentes méthodes : la méthode de l'inverse de la distance, l'utilisation de *splines*, les différentes formes du krigeage (universel ou ordinaire) ou encore, d'une façon plus générale, les processus gaussiens. Dans le cadre de ce mémoire, nous nous attarderons sur les méthodes associées aux processus gaussiens.

Ces derniers se rencontrent dans de nombreux domaines [Chen et al., 2014] [Schulz et al., 2018] [Quiñonero-Candela and Rasmussen, 2005], avec principalement deux applications : la régression et la classification. Seule la première des deux est importante dans le cadre de ce mémoire car elle permet notamment d'interpoler des données spatiales ou bien de réaliser des prédictions.

Le principe d'une régression est de déterminer une fonction décrivant le passage entre un *input* et un *output*. Dans certains cas, la fonction est considérée comme connue (par exemple : les régressions linéaires). Le processus gaussien permet de traiter les cas pour lesquels la forme de la fonction est considérée inconnue et difficile à déterminer de façon analytique. Dans ces cas, l'acquisition de données est coûteuse ou techniquement difficile [Schulz et al., 2018].

L'utilisation de processus gaussiens lors de la réalisation d'une régression peut avoir deux objectifs : (1) extrapoler un ensemble de trajectoires possibles ou (2) déterminer un ensemble d'interpolations possibles [Rasmussen and Williams, 2006]. Les méthodes associées aux régressions grâce aux processus gaussiens sont dérivées des approches bayésiennes non paramétriques [Gershman and Blei, 2012]. Cette approche permet de déterminer un modèle unique qui adapte sa complexité en fonction des données. Il est ainsi possible d'éviter le test et la comparaison de nombreux modèles différents, ou l'emploi de trop de paramètres, avec les problèmes bien connus de la surdétermination des modèles (*overfitting*) [Gershman and Blei, 2012].

Les modèles bayésiens définissent un processus latent², qui détermine une distribution *a priori* des états possibles du système [Gershman and Blei, 2012]. Le théorème de Bayes permet de corriger cette distribution en utilisant les observations. La distribution corrigée est le *posterior*. La figure 1.4 illustre ce principe. Les points de données (en bleu) permettent de déterminer la forme de la fonction en vert qui définit ici le processus latent. Cette fonction n'est pas associée à une forme analytique (par exemple, a priori ce n'est pas un polynôme) et c'est grâce aux observations qu'il est possible de la déchiffrer.

L'inférence bayésienne permet de réaliser ce calcul pour des fonctions complexes avec un nombre important de points. Ces points sont des observations de terrain, ce qui implique qu'elles sont contaminées par du bruit. Les modèles basés sur les processus gaussiens permettent de prendre en compte ce bruit et de fournir une distribution normale en chacun des points.

Dans le cadre de données associées au climat (température, précipitations, ...), l'intégration de la covariance entre les valeurs peut permettre l'amélioration de la performance du modèle d'interpolation, ce qui est également possible dans les processus gaussiens.

2. Non-observé

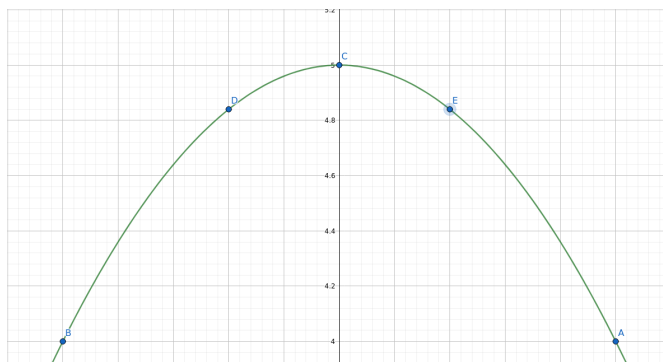


FIGURE 1.4 – Fonction depuis une équation $f(x) = 5 - x^2$

L'utilisation d'un modèle non paramétrique permet d'éviter la spécification, *a priori*, de la complexité de la fonction latente. Cette approche est contraire à ce qui est fait dans les régressions classiques. Cet avantage nécessite toutefois de spécifier une fonction de covariance permettant de quantifier le lien entre différentes observations.

Lors d'une régression, les *inputs* sont les points de données fournis par les archives. Ces points de données sont ensuite liés grâce à la fonction de covariance.

Lors d'une reconstruction spatiale, les *inputs* sont des points disposant d'une position dans l'espace. Il est généralement admis que deux observations proches sont davantage corrélées que des observations lointaines. Cette hypothèse permet d'utiliser des équations liant la corrélation entre les observations et la distance les séparant.

L'utilisation de processus gaussiens peut ainsi être considérée comme une extension de la méthode utilisée pour déterminer l'équation de la figure 1.4. Ils représentent une densité de probabilités de fonctions qui peuvent être modifiées grâce à des données empiriques. Diverses fonctions sont ainsi possibles *a priori* pour une même moyenne et déviation standard (cas théorique présenté à la figure 1.5).

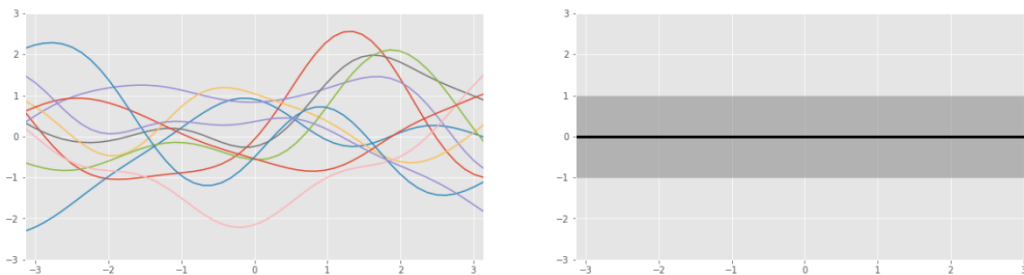


FIGURE 1.5 – *Prior* dans le cas d'un processus gaussien (moyenne = 0, déviation = 1) [Knagg, 2019]

Dans le cadre de l'étude du paléoclimat, des résultats de la fonction générale inconnue "paléoclimat" sont observables grâce aux archives. Il est donc possible de contraindre les fonctions aléatoires *a priori* (partie gauche de la figure 1.5) à passer par les valeurs empiriques afin d'améliorer l'approximation de la fonction inconnue (partie gauche de la figure 1.6). Cette méthode sera utilisée pour combler les trous dans les séries temporelles ainsi que pour interpoler spatialement les données.

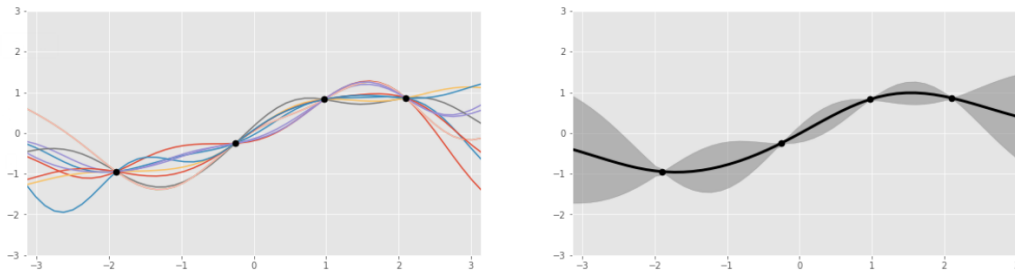


FIGURE 1.6 – Restriction des fonctions grâce aux données [Knagg, 2019]

Les processus gaussiens étant des modèles probabilistes, le *posterior* est une distribution de probabilités de fonctions latentes. Il est donc possible de calculer différentes possibilités sur base d'une même fonction de covariance. Suite à l'utilisation des processus gaussiens, ces différentes possibilités ne sont pas tirées au hasard. Elles sont des possibilités pour la fonction latente. Le tirage d'un grand nombre de ces valeurs possibles permet ensuite de se rapprocher de la forme de la fonction latente.

Assimilation de données

Les méthodes d'assimilation de données permettent de combiner des modèles climatiques avec des données empiriques. Les trajectoires des modèles peuvent ainsi être contraintes par les données. L'utilisation du modèle permet, quant à elle, de garantir un cadre physique pour la reconstruction. Une des méthodes possibles est l'ajout d'un terme de forçage artificiel aux équations prédictives afin de rapprocher les états simulés des observations (méthodes de *nudging*) [Fang and Li, 2016]. Cet ajustement est plus ou moins simple en fonction du nombre de degrés de liberté du modèle. Ce type de reconstruction est extrêmement dépendant des modèles utilisés et de leurs biais. Ces biais peuvent provenir de différents facteurs, notamment les processus physiques méconnus non intégrés ou bien les choix de paramétrisation d'autres éléments. Les résultats sont également dépendants de la distribution spatiale des archives. L'apport de cette méthode pour les zones éloignées des données empiriques est en effet assez faible. Le résultat final est également dépendant des conditions initiales fournies au modèle et de la technique d'assimilation. De plus, une vérification est compliquée à mettre en place car aucune donnée instrumentale n'est disponible afin de réaliser une comparaison. De longues périodes de reconstructions sont également coûteuses en temps de calcul, ce qui limite leur utilisation [Fang and Li, 2016].

Processus gaussiens et modèle *LOVECLIM*

Ce mémoire propose une troisième approche qui combine les avantages de l'interpolation et de l'assimilation de données. Il est ainsi possible d'exploiter le cadre physique fourni par un modèle (\sim méthodes d'assimilation de données) tout en utilisant un maximum d'informations des archives. Cette combinaison se réalise grâce aux processus gaussiens. L'apport de ce travail est d'utiliser la structure de covariance du modèle *LOVECLIM* à la place d'une fonction simple de la distance. Comme il s'agit toujours d'une interpolation, les observations déterminent bien les reconstructions de températures, mais l'interpolation tiendra compte de la structure spatiale des modes de variabilités climatiques simulés avec le modèle *LOVECLIM* (figure 1.6). L'approche conserve l'avantage apporté par la rapidité d'exécution des méthodes d'interpolation classiques. La structure de covariance du modèle, quant à elle, conserve le cadre physique fourni par le modèle.

La structure de corrélation sert de squelette à l'interpolation dans laquelle les informations des archives s'intègrent : la méthode conserve le meilleur des deux mondes (table 1.1).

TABLE 1.1 – Résumé des caractéristiques des méthodes présentées

	Assimilation de données	Interpolation	Processus gaussiens et structure de covariance
Processus physiques représentés	++	-	++
Information des archives utilisées	+	++	++
Temps de calcul	-	++	+
Compétences techniques requises	-	+	+

Ce mémoire s'attarde donc sur la comparaison des méthodes d'interpolation qui utilisent une forme paramétrisée de la covariance avec une méthode d'interpolation qui utilise une structure de covariance fournie par un modèle.

Chapitre 2

Matériel et méthode

2.1 Matériel

2.1.1 Modèle *LOVECLIM*

Le modèle climatique *LOVECLIM* est un modèle de complexité intermédiaire compilant cinq composantes couplées (atmosphère, océans et glace de mer, terres émergées avec végétation, calottes polaires, icebergs et cycle du carbone) (figure 2.1) [Goosse et al., 2010]. Il est également intéressant d'utiliser ce modèle car, étant de complexité intermédiaire, la manipulation des données peut se réaliser facilement sur un ordinateur personnel (64 cellules en longitude et 32 cellules en latitude).

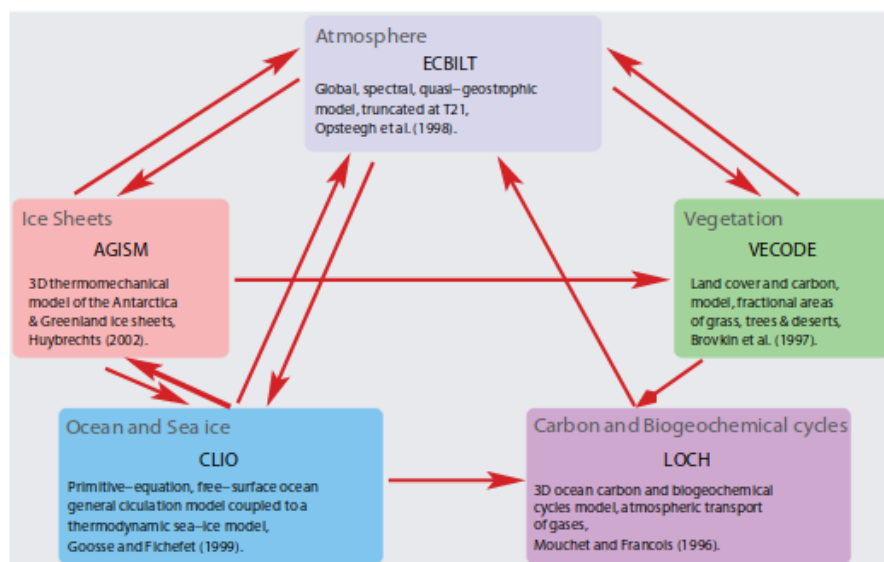


FIGURE 2.1 – Modèle *LOVECLIM* : Composantes et interactions [Goosse et al., 2010]

Ce modèle n'est toutefois pas parfait. Il est important de préciser les biais auxquels notre reconstruction sera soumise, notamment au niveau des températures. En effet, le modèle surestime de façon importante ces dernières aux basses latitudes. La sensibilité climatique¹ est de 1.9°C après 1000 ans [Goosse et al., 2010]. Cette valeur est assez faible, notamment lors de la comparaison avec les valeurs de modèle *GCM* [Goosse et al., 2010].

1. Augmentation des températures de surface lorsque les concentrations en CO_2 sont doublées

Deux simulations d'un ensemble de températures mensuelles pour chaque cellule durant 2000 ans ont été réalisées. Une première simulation n'est pas forcée et les paramètres d'entrées sont ceux (déterminés arbitrairement) de l'année 1400. La deuxième simulation est forcée. Un grand nombre de forçages peuvent être pris en compte par le modèle. Les forçages naturels, au nombre de trois, sont : l'activité volcanique, les variations du cycle solaire et les changements des paramètres orbitaux [Driesschaert, 2005]. Les forçages anthropiques sont également au nombre de trois. Il s'agit des changements dans les concentrations de gaz à effet de serre, des variations dans la charge des aérosols sulfatés et de la déforestation [Driesschaert, 2005].

Les données mensuelles sont tout d'abord moyennées annuellement. Les années 1800 à 2000 sont supprimées et les données des années 500 à 1800 sont standardisées. En effet, comme visible sur la figure 2.2, le modèle est capable de reconstruire un réchauffement anthropogénique suite à l'évolution du forçage associé aux gaz à effet de serre, ce qui aura pour effet de tronquer la corrélation entre les cellules. De plus, nous nous intéressons aux variations naturelles du climat sur le long terme. L'ajout de l'influence des forçages associés aux activités anthropiques dans la structure de corrélation ne semble donc pas nécessaire.

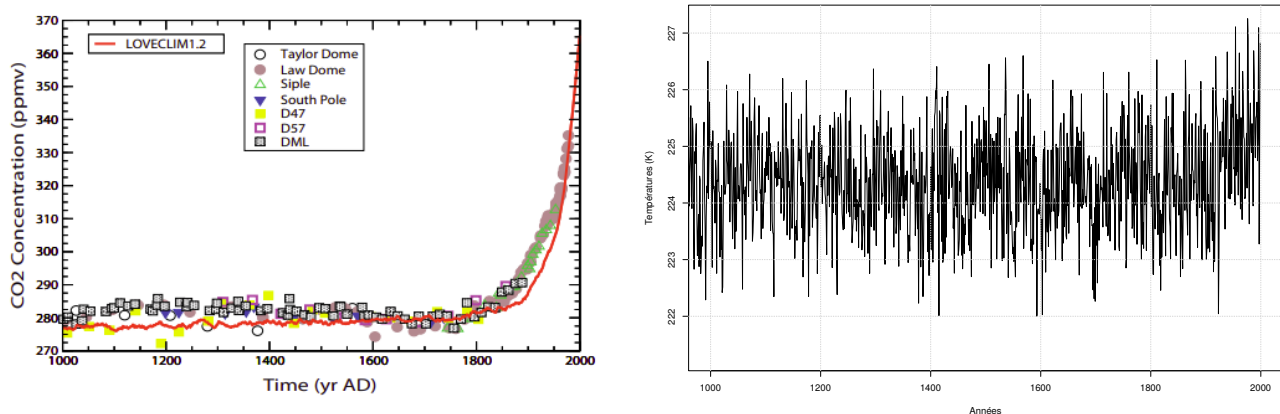


FIGURE 2.2 – Évolution des concentrations en CO_2 et de la température [Goosse et al., 2010]

2.1.2 Archives

La base de données regroupant les différentes archives est disponible en [open access](#) et gérée par le programme PAGES2K [Kilbourne et al., 2017]. Cette base de données regroupe les résultats de nombreuses études ayant pour objectif de faire un lien entre les archives, les *proxies* et éventuellement l'évolution de la température au cours du temps.

La base de données regroupe 657 études. Toutefois, pour la plupart des archives, la base de données n'établit pas le lien entre le *proxy* et la température. Ainsi, pour les largeurs de cernes d'arbres par exemple, une estimation de la température n'est pas fournie. Il est possible de réaliser ce lien entre observation et température, mais ce travail n'est pas l'objectif du mémoire et les données ne correspondant pas à des températures sont donc ignorées. Les seules archives restantes sont les sédiments marins et les sédiments lacustres (89).

Ces archives sont réparties sur le globe (figure 2.3) et sont les données utilisées lors de la partie méthodologie. Tout comme les données de *LOVECLIM*, les valeurs des années supérieures à 1800

sont ignorées et les données des années restantes sont standardisées.

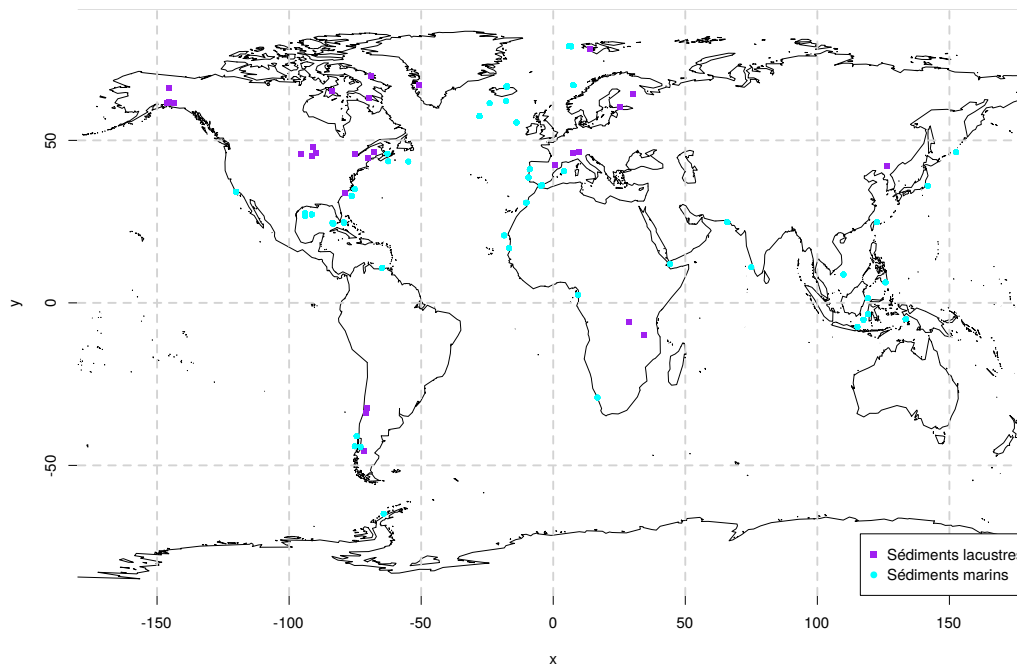


FIGURE 2.3 – Localisation des archives climatiques

2.2 Méthode

Deux défis seront abordés dans cette partie. Le premier défi est lié aux séries temporelles des archives. Ces dernières ne sont pas complètes sur l'intervalle de temps considéré (années 500 à 1800). Ce manque de données sur une série temporelle affecte la contrainte du *prior* par les données lors de l'interpolation spatiale. Les périodes sans données seront donc remplies.

Le deuxième défi consiste à combiner la structure de corrélation fournie par *LOVECLIM* et les données des archives.

Ces données seront également utilisées pour réaliser une interpolation spatiale de référence. Cette interpolation utilise une paramétrisation de la corrélation en fonction de la distance.

2.2.1 Processus gaussiens et séries temporelles

Le modèle statistique utilisé dans cette partie est défini par l'équation :

$$y = f(t) + \epsilon(t) + \zeta(t), \quad (2.1)$$

avec y : la valeur observée, $f(x)$: le climat réel, ϵ : l'erreur de l'observation dépendante à la fois de la façon dont les archives retiennent l'information et des erreurs lors de la conversion en température et ζ la variabilité à haute fréquence. Ce dernier terme est un ajout par rapport à la formulation de [Rasmussen and Williams, 2006]. Dans notre cas, l'intérêt se porte sur le processus climatique à basse fréquence et non pas sur la variabilité interannuelle. Ces deux éléments sont donc séparés.

Les différentes observations d'une même série sont bruitées. Chacun des points d'une série temporelle peut alors être associé à une distribution gaussienne. La moyenne correspond à la valeur fournie par les données brutes et la variance est déterminée par le bruit. Nous faisons ici l'hypothèse que cette valeur de bruit est équivalente pour toutes les valeurs connues d'une série. Cela nous évitera de compliquer les calculs suite à l'augmentation du nombre de paramètres libres.

Séries temporelles

Dans un premier temps, nous considérons une seule série temporelle de valeurs par archives. Ces séries temporelles ne sont pas complètes sur l'intervalle de temps considéré. Ces séries temporelles doivent donc être remplies grâce aux processus gaussiens afin de disposer d'une estimation de la température (avec une variance associée) pour chaque année. L'utilisation des processus gaussiens pour une régression nécessite une spécification de la forme de la covariance entre les observations (équation 2.2). Il n'est pas possible, suite à la forme de la fonction de covariance choisie ici, de fournir une estimation de la variance de $\epsilon(t)$ et de $\zeta(t)$. La variance que nous mesurons sera donc la somme de ces deux variances.

$$C(t_1; t_2) = \exp\left(-\frac{\|t_1; t_2\|}{L}\right) \quad (2.2)$$

Cette fonction fait appel à la distance entre les observations $\|t_1; t_2\|$ et à un paramètre L . Ce paramètre est la longueur d'échelle et s'interprète comme le rayon d'influence d'une valeur. Ce paramètre est *a priori* inconnu mais il est possible de l'estimer par vraisemblance marginale (équation 2.3).

$$\log p(\mathbf{y}|X) = -\frac{1}{2}\mathbf{y}^\top (K + \sigma_n^2 I)^{-1} \mathbf{y} - \frac{1}{2} \log |K + \sigma_n^2 I| - \frac{n}{2} \log 2\pi \quad (2.3)$$

où $|K_y|$ est le déterminant de la matrice $K_y = K_f + \sigma_n^2 I$ (matrice de corrélation pour les données bruitées). Une seule des trois parties de cette équation, $-\frac{1}{2}\mathbf{y}^\top K_y^{-1} \mathbf{y}$, est dépendante des données tandis que la deuxième partie, $\frac{1}{2} \log |K_y|$, est un terme de pénalité associé à la complexité de la matrice K_y . La troisième partie, $\frac{n}{2} \log 2\pi$ est une constante de normalisation. Ce calcul est appliqué à l'ensemble des séries temporelles disponibles et une moyenne des valeurs L est utilisée pour la suite.

C'est à ce stade que le théorème de Bayes est utilisé pour mettre à jour le *prior* défini par le processus gaussien, ce qui nous permet d'en estimer la moyenne (équation 2.4) et la covariance (équation 2.5) *a posteriori* :

$$\bar{\mathbf{f}}_* \triangleq \mathbb{E}[\mathbf{f}_* | X, \mathbf{y}, X_*] = K(X_*, X) [K(X, X) + \sigma_n^2 I]^{-1} \mathbf{y} \quad (2.4)$$

$$\text{cov}(\mathbf{f}_*) = K(X_*, X_*) - K(X_*, X) [K(X, X) + \sigma_n^2 I]^{-1} K(X, X_*) \quad (2.5)$$

L'indicateur $K(\cdot, \cdot)$ fait référence à la covariance entre les éléments de la parenthèse, X_* représente les points où une évaluation de la fonction est nécessaire (dans le cas du remplissage des séries, le but est d'avoir une valeur par année donc X_* est compris entre 500 et 1800), X est l'ensemble des localisations des points dont la valeur est connue (il varie donc en fonction de la série considérée) et y représente les valeurs de la série connue. L'inversion de la matrice $[K(X, X) + \sigma_n^2 I]$ est réalisée

par la décomposition de Cholesky, beaucoup plus rapide que l'inversion matricielle classique. Le terme σ_n^2 reste toutefois inconnu et il est nécessaire de le spécifier *a priori* lors d'une application classique des processus gaussiens. Sa valeur sera fixée à 0.7 dans ce travail.

L'équation 2.4 permet de déterminer un vecteur moyen de dimension $[1300 \times 1]$ qui représente la moyenne de l'ensemble des états possibles du système tandis que l'équation 2.5 détermine une matrice $[1300 \times 1300]$ reprenant les écarts par rapport à cette moyenne. Grâce à ces deux éléments, il est possible d'échantillonner la distribution normale, définie par cette moyenne et cette covariance, afin de disposer d'un ensemble de reconstructions probables de la température au cours des 1300 années. Ce tirage de reconstructions est réalisé 100 fois par échantillon. Ces tirages permettent ensuite l'estimation de la variance de chacun des points.

La réalisation de ces tirages est décrite dans [Hofert, 2013] et nous en reprenons le principe ici. La distribution normale multivariée est définie comme :

$$X = \mu + AZ \tag{2.6}$$

L'élément $\mathbf{Z} = (Z_1, \dots, Z_k)$ est un vecteur aléatoire de dimension k . Chacun des éléments de ce vecteur est une variable aléatoire normale standard indépendante.

L'élément A est une matrice de dimension (d, k) et μ est le vecteur des valeurs moyennes. La matrice de covariance de X peut être définie comme $\Sigma = AA^t$. Il est possible d'exprimer la distribution de X grâce à un vecteur moyen et une matrice de covariance. Il est donc possible d'écrire $X \sim \mathcal{N}(\mu, \Sigma)$. La matrice de covariance Σ , positive définie, peut être décomposée en :

$$\Sigma = BB^T \tag{2.7}$$

Cette matrice est le facteur de Choleski de la décomposition de Choleski. Cette décomposition est particulièrement efficace pour solutionner des systèmes d'équations linéaires. Il est ensuite possible d'utiliser l'équation 2.6 en remplaçant A par la valeur de B .

L'approche classique des processus gaussiens nécessite donc de spécifier la variance des observations avant le calcul de la longueur d'échelle. Cela pose problème lorsque cette variance n'est pas précisée par le fournisseur de données. L'estimation de ce bruit serait toutefois possible pour un L donné. Ce paramètre peut être vu comme la persistance d'une valeur dans le temps. Cette persistance peut être estimée grâce à l'autocorrélation de la série temporelle considérée. Cette série n'est pas disponible pour les séries de données des archives (le problème ne serait pas présent si cette série était disponible) mais est disponible grâce au modèle *LOVECLIM*. La diminution de l'autocorrélation d'une série temporelle avec le temps peut être modélisée comme :

$$C(t_1; t_2) = c + (d - c) \exp\left(-\frac{\|t_1; t_2\|}{\tau}\right) \tag{2.8}$$

Cette équation et l'équation 2.2 sont relativement similaires. L'autocorrélation d'une série temporelle *LOVECLIM* est ainsi dépendante du paramètre c : la valeur minimale atteinte, du paramètre d la valeur maximale atteinte par l'autocorrélation et du paramètre τ qui détermine la pente de la décroissance. Les valeurs minimales (c) et maximales (d) sont fixées aux valeurs de 0 et 1, respectivement. C'est à ce stade que l'on peut justifier le terme $\zeta(t)$ dans l'équation 2.1. En effet, l'utilisation des données annuelles pour le calcul de l'autocorrélation donne des résultats pollués par la variabilité interannuelle. L'autocorrélation diminue de façon extrêmement rapide et devient

même, pour certaines cellules, inférieure au seuil de significativité² après un an :

$$\frac{(1 + 0.95)/2}{\sqrt{\sum_{i=1}^i x_i}} \quad (2.9)$$

L'autocorrélation des séries temporelles n'est donc pas significative et n'est pas représentative des modes de variabilités plus lents. En réalité, ce qui est donné comme résultat correspond à la variabilité à haute fréquence qui a son utilité pour étudier des phénomènes tels que *El Nino* ou l'oscillation nord-atlantique. Néanmoins, ce qui est recherché dans ce mémoire concerne la variabilité à basse fréquence et les phénomènes de variation de la température sur le long terme. Cette séparation peut être réalisée grâce à un filtre. Ce filtre est une moyenne mobile dont la fenêtre temporelle est variable. Cette valeur est arbitraire. L'utilisation de ce filtre génère de l'autocorrélation mais supprime une grande partie de l'information disponible dans les séries. Cette méthode permet néanmoins de séparer, en partie, la variabilité à haute fréquence et la variabilité à basse fréquence. A ce stade, la période de filtrage est fixée à 100 ans de façon arbitraire.

Ce calcul n'est toutefois pas appliqué tel quel à toutes les valeurs d'autocorrélation calculées à tous les délais³, un seuil de significativité est mis en place. Les valeurs très lointaines sans influence sont plus nombreuses que les valeurs proches et font varier considérablement le résultat (figure 2.4). L'autocorrélation a également tendance à osciller dès que le délai devient trop important. Les valeurs d'autocorrélation sous le seuil calculé par l'équation 2.9 ne sont pas considérées dans le calcul.

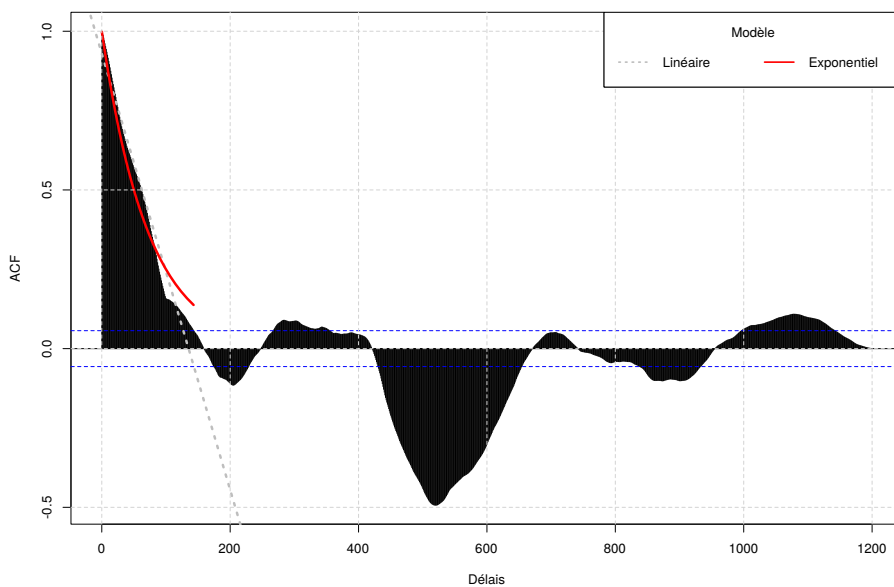


FIGURE 2.4 – Autocorrélation *LOVECLIM*, exponentielle et seuils de significativité

Cette utilisation de *LOVECLIM* permet d'estimer le paramètre τ . Il est alors possible d'utiliser la vraisemblance marginale (équation 2.3) pour estimer le paramètre σ_n^2 . Le théorème de Bayes peut de nouveau être appliqué pour estimer la moyenne et la covariance. Les tirages permettent une estimation de la variance de chacune des années.

2. Ce seuil correspond à des valeurs d'autocorrélation pour une série générée de façon aléatoire

3. Un délai est l'écart de temps entre la valeur au temps 1 et le temps considéré

2.2.2 Processus gaussiens et interpolation spatiale

L'interpolation spatiale grâce aux processus gaussiens est relativement directe car toutes les équations ont déjà été présentées (équations 2.4 et 2.5). Les séries temporelles sont complètes à ce stade. Il est donc possible de réaliser une interpolation spatiale qui exploite la corrélation entre les cellules.

Il est tout d'abord nécessaire de réaliser des étapes de prétraitement entre les localisations des cellules *LOVECLIM* et la localisation des *proxies*. En effet, les résultats seront présentés sous la forme d'une grille ayant la même disposition que le modèle, ce qui nécessite de lier les données empiriques au point de grille le plus proche (figure 2.5). Les données sont supposées être représentatives d'une certaine région et, en faisant cette hypothèse, le déplacement de l'information dans le point de grille le plus proche n'est donc *a priori* pas problématique.

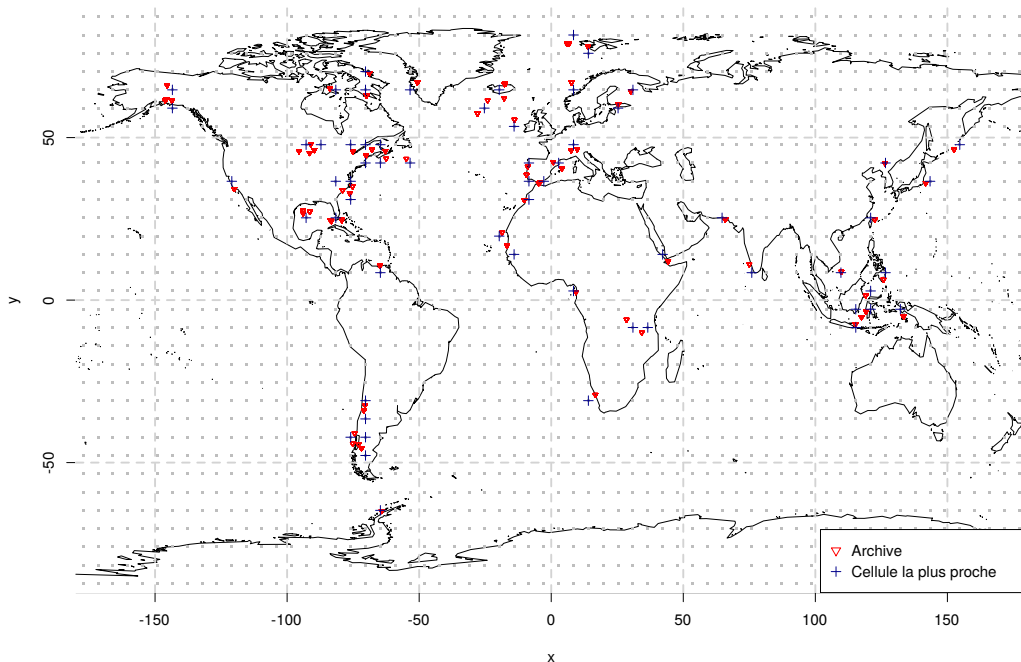


FIGURE 2.5 – Localisation des points de grille *LOVECLIM* et cellules les plus proches des données empiriques

Ces cellules disposant d'une information empirique seront notées X et les cellules sans information empirique X_* . A ce stade, il serait possible de paramétrer la corrélation en fonction de la distance qui sépare les points de grille. Cela revient à utiliser une forme de la covariance comme l'équation 2.2, non plus dépendante de la distance temporelle, mais bien de la distance dans l'espace. Cette approche sera également présentée. Le rayon d'influence de chacune des observations ne sera pas optimisé et sera fixé à trois valeurs différentes : 500, 1000 et 1500.

Dans le cadre de l'étude d'un champ global de températures, cette hypothèse n'est pas réaliste. En effet, certaines régions du monde présentent des variations de températures fortement corrélées, même à grande distance. C'est le cas dans le Sahara ou en Sibérie. A l'inverse, cette corrélation peut rapidement diminuer dans l'espace. C'est ce qui est observé au niveau des chaînes de montagnes. La dynamique des températures peut y être radicalement différente en fonction du versant étudié.

Ce lien entre les différentes localisations du monde ne peut donc pas se paramétrer en fonction de la distance. Il est toutefois nécessaire de disposer d’une information sur les liens entre cellules lors d’une interpolation géostatistique. Cette information sera fournie par le modèle *LOVECLIM*. Les différentes cellules sont liées grâce à la physique intégrée dans le modèle. Cette information peut donc être utilisée comme le serait une matrice de corrélation calculée en fonction de la distance.

Cette corrélation entre les cellules est donc calculée sur base des séries temporelles simulées par le modèle. La matrice $K(X, X)$ est donc la matrice de corrélation entre les cellules disposant d’archives. Si 50 archives sont disponibles, sa taille sera 50×50 . La matrice $K(X_*, X_*)$ correspond à la corrélation entre l’ensemble des cellules du modèle et sa dimension est de 2048×2048 . Les matrices $K(X_*, X)$ et $K(X, X_*)$ correspondent à la corrélation entre les cellules avec une information et les cellules sans information empirique. Leurs dimensions sont respectivement de 2048×50 et 50×2048 .

La matrice diagonale du bruit (σ_n^2) des observations est également connue. Les tirages temporels représentant un écartement des valeurs par rapport à la moyenne, il est possible d’estimer la variance de la distribution gaussienne en chacun des points. Cette estimation se réalise grâce à la variance de 100 tirages. L’optimisation du bruit par la vraisemblance marginale n’est donc plus nécessaire.

L’insertion de ces matrices dans l’équation 2.4 (moyenne) et dans l’équation 2.5 (covariance) permet, tout comme pour combler les séries temporelles, de disposer d’une moyenne de l’état et de la covariance associée (\sim variations possibles) sur tout le globe.

Le tirage d’un grand nombre de scénarios a déjà permis d’estimer la variance de la distribution en chacun des points d’une série temporelle. Les tirages permettent alors d’estimer la variance de chacune des cellules de la reconstruction.

2.2.3 Analyse en composantes principales

Une analyse en composantes principales est une méthode statistique permettant d’analyser la variabilité d’un champ quelconque. Cette méthode permet de résumer l’information des différentes structures spatiales et temporelles présentes dans les données [Bjornsson and Venegas, 1997]. Les résultats de cette analyse sont appelés des modes et ont des poids différents dans l’explication de la variance du jeu de données. Cette méthode sera appliquée à la reconstruction temporelle.

La première étape est de créer une grande matrice F composée d’un nombre de lignes équivalant au nombre d’années de la reconstruction et d’un nombre de colonnes équivalant au nombre de stations (figure 2.6). La seconde étape est de calculer la matrice de covariance R de F tel que $R = F^t F$.

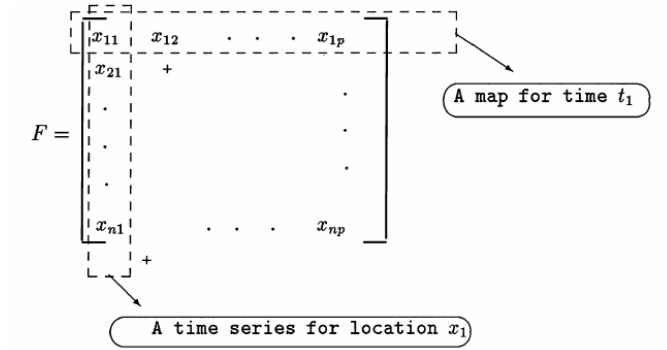


FIGURE 2.6 – Matrice de données pour l’ACP [Bjornsson and Venegas, 1997]

Cette matrice sert à résoudre le problème des valeurs propres selon :

$$RC = C\Lambda \tag{2.10}$$

avec Λ la matrice diagonale contenant les valeurs propres λ_i de R . Les colonnes de la matrice C contiennent les vecteurs propres. Chacun de ces vecteurs est un mode de variabilité. Ces modes expliquent une partie de la variance totale observée [Gould, 1967]. Seules les trois premières composantes seront utilisées dans le cadre de ce travail car elles reprennent généralement une grande part de l’information.

Chapitre 3

Résultats

3.1 Séries temporelles

3.1.1 Remplissage

Le premier objectif de ce travail était de combler des séries temporelles éparses grâce aux processus gaussiens. Le lien entre les observations est paramétré comme une exponentielle décroissante et c'est la méthode utilisée pour déterminer le facteur d'échelle qui est variable. Deux méthodes sont utilisées, la première est de chercher la valeur de ce paramètre qui optimise la vraisemblance marginale ; la seconde est d'utiliser l'autocorrélation des données lissées (100 ans) du modèle *LOVECLIM*.

Les valeurs des paramètres τ et L sont ainsi relativement différentes (figure 3.1). L'ajout des forçages ou non dans la simulation *LOVECLIM* n'influençant que légèrement l'autocorrélation des séries temporelles, les valeurs de τ sont très proches (55 et 67).

Le modèle faisant appel à l'optimisation de la vraisemblance marginale est la méthode présentant la plus grande valeur moyenne du paramètre de décroissance (311). Cette valeur L moyenne est toutefois calculée sur un grand ensemble de valeurs différentes. Les valeurs L occupent en effet l'ensemble des valeurs comprises dans les bornes de l'optimisation (entre 0 et 1000). La moitié des valeurs présente également une étendue importante, ce qui n'est pas le cas lors de l'utilisation des données *LOVECLIM*. L'ensemble des valeurs de cette méthode est plus restreint.

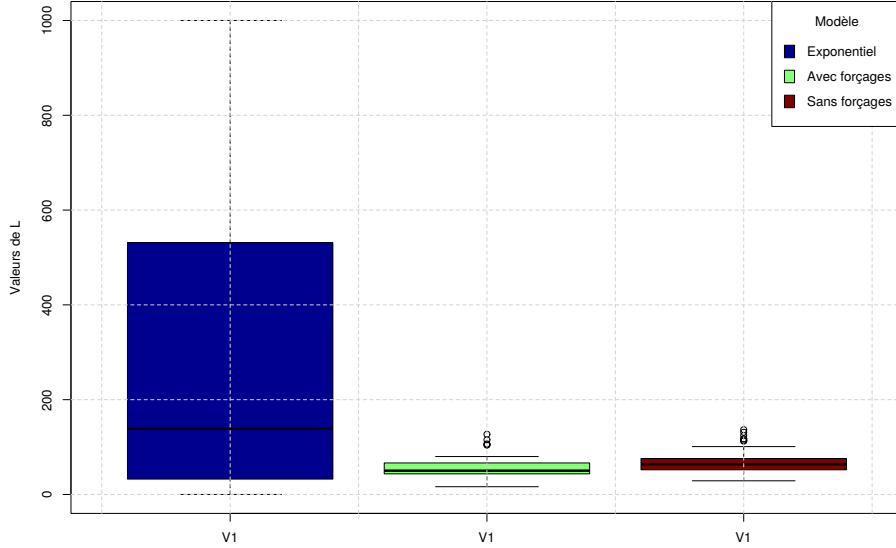


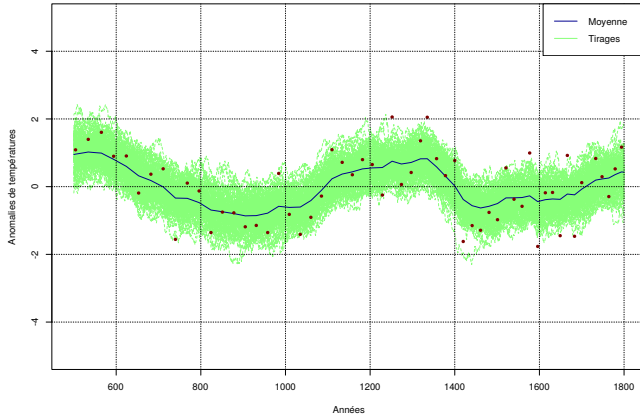
FIGURE 3.1 – Valeurs de L en fonction du modèle

L'utilisation de la valeur moyenne des L fournie par optimisation de la vraisemblance marginale (311) fournit un remplissage de séries temporelles présentant des variations à très basses fréquences. Ce comportement est typique de l'utilisation des processus gaussiens. Lorsque la valeur du bruit est élevée, il est possible d'optimiser une grande valeur L . Le modèle est en effet beaucoup plus libre dans son passage proche des points de données. Une partie de la distribution gaussienne en chacun des points peut être considérée comme aberrante mais une partie de la distribution, grâce au bruit élevé, n'est pas considérée comme aberrante. C'est cette partie non aberrante qui permet le passage lisse du modèle lors de l'utilisation d'une longueur d'échelle élevée.

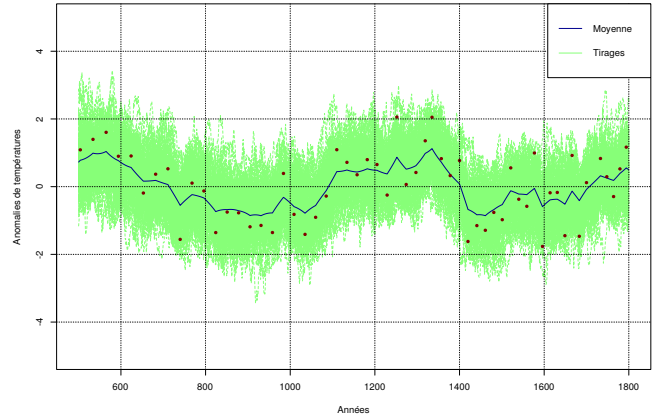
L'importance d'une observation particulière est réduite si elle est entourée d'autres observations dans le rayon L . L'utilisation d'un modèle exponentiel avec la valeur τ présente un autre remplissage (figure 3.2a). La variabilité reconstruite se situe dans une gamme supérieure de fréquence. La différence étant très faible entre le modèle forcé et non forcé, seul le remplissage grâce à la valeur τ du modèle forcé est présentée. La reconstruction temporelle est marquée par des changements abrupts de la pente de la courbe moyenne au cours du temps. Ces ruptures de la courbe moyenne correspondent à des changements rapides et abrupts de températures entre plusieurs années.

Il est également possible de constater que la variance des tirages est nettement plus faible sur la figure 3.2a que sur la figure 3.2b. Cette différence de variance des tirages est notamment causée par une valeur σ_n^2 inférieure sur les séries temporelles lors de l'utilisation de la valeur L . Cette valeur plus faible de σ_n^2 a un impact sur la variance. Les différents tirages doivent en effet passer au plus proche des points de données, ce qui diminue la variance.

Une partie des changements abrupts reconstruits sont probablement causés par l'utilisation d'un modèle exponentiel. L'influence du choix d'un modèle par rapport à un autre sera abordé dans la partie discussion.



(a) Modèle exponentiel

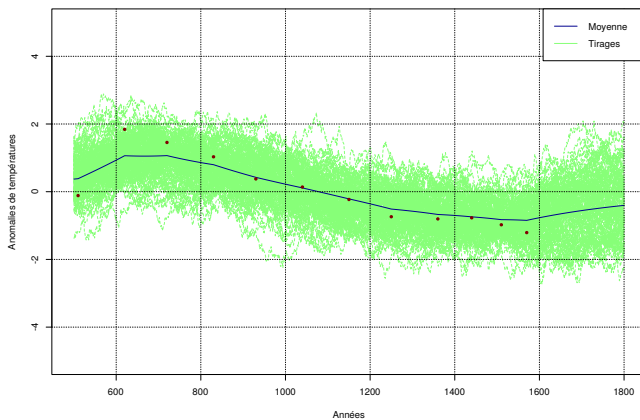


(b) Modèle forcé

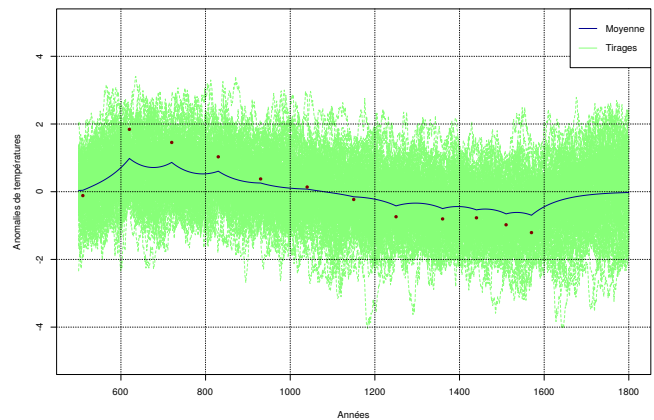
FIGURE 3.2 – Série temporelle remplie

Pour certaines séries temporelles, l'utilisation de la valeur L déterminée par l'autocorrélation crée un remplissage incohérent avec un processus réel de variation des températures (figure 3.3b). Les variations de la courbe moyenne présentent en effet une oscillation à proximité des observations. Cette oscillation est liée à la faible valeur de la longueur d'échelle. Ce problème n'est pas présent sur les tirages.

Lors de l'éloignement d'un point, l'influence d'une valeur diminue très rapidement. La valeur suivante la plus proche étant hors du rayon d'influence L , elle ne contraint pas le modèle. La contrainte de cette deuxième valeur n'intervient que lorsque l'influence de la première valeur est quasiment nulle, la contrainte étant alors uniquement causée par la seconde valeur ce qui crée une oscillation. Sur la série temporelle choisie, le modèle ne prend ainsi que légèrement en compte la tendance linéaire des données. L'utilisation de cette série oscillante lors de la reconstruction spatiale introduira un biais important en occultant une partie du refroidissement graduel de la région. Il est toutefois important de noter que lors de l'utilisation de la valeur L élevée, le refroidissement est sous-estimé suite à la présence d'une valeur proche de 0 vers 550.



(a) Remplissage cohérent : modèle exponentiel



(b) Remplissage incohérent : modèle *LOVECLIM*

FIGURE 3.3 – Séries temporelles : incohérences

Le remplissage des séries temporelles est également influencé par la valeur du bruit lorsque la méthode choisie fait appel à l'autocorrélation. Les observations sont ainsi plus ou moins contraignantes en fonction de la valeur de σ_n^2 . Cette contrainte est réduite lorsque le bruit augmente. Ces valeurs de bruit sont relativement élevées lors de leur optimisation (figure 3.4). La contrainte par les observations est donc finalement relativement faible, ce qui n'empêche toutefois pas le remplissage des séries de sembler pertinent.

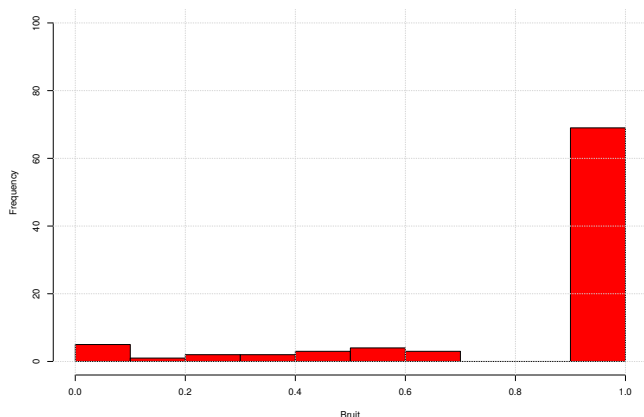


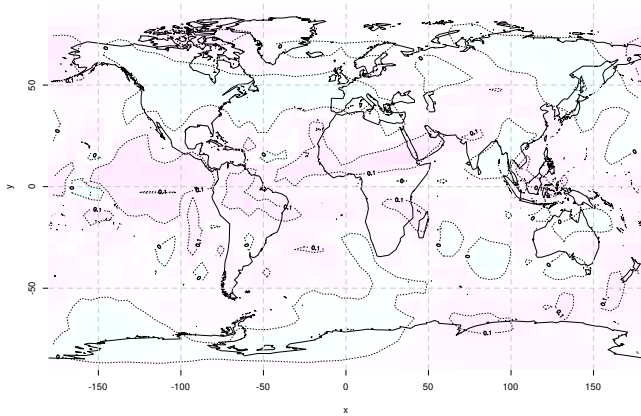
FIGURE 3.4 – Valeurs du bruit optimisé par vraisemblance marginale (modèle forcé)

3.2 Champ spatial

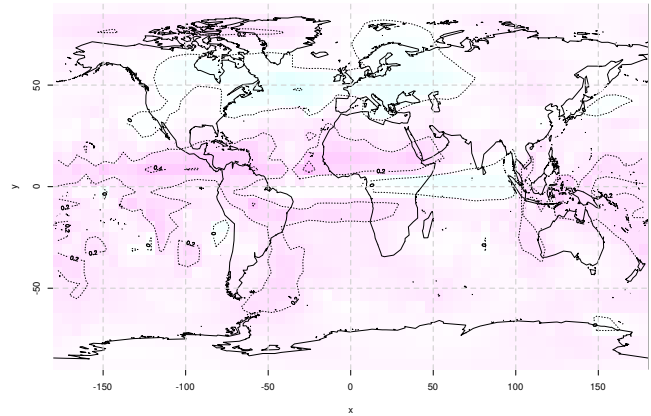
3.2.1 Structures de corrélation : *LOVECLIM*

L'interpolation spatiale est influencée par deux éléments : la corrélation entre chaque cellule et la confiance accordée à chacune des observations. Le premier de ces deux éléments n'est pas variable au cours de la reconstruction. Il présente toutefois des différences en fonction du choix de la forme du lien entre les cellules (corrélation en exponentielle décroissante ou tirée d'un modèle). Ces différences sont logiquement importantes entre une structure exponentielle décroissante et une structure fournie par *LOVECLIM*. Les différences entre une structure prenant en compte les forçages et une structure sans forçages sont quant à elles variables (figure 3.5).

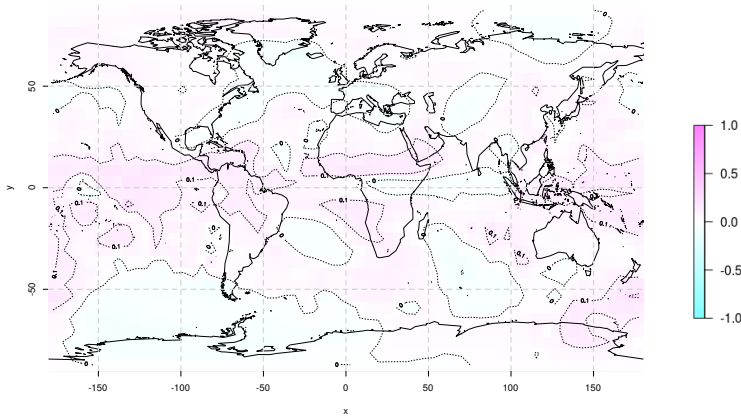
Les différences de corrélation sont fortement dépendantes de la localisation de la cellule. Les différences globales par rapport aux cellules polaires, hémisphère nord ou sud, sont ainsi limitées (entre -0.1 et 0.1). Ces différences sont néanmoins marquées pour les cellules se situant dans la région intertropicale. La différence de corrélation entre le modèle forcé et le modèle non forcé se situe entre -0.2 et 0.4 dans ces régions. Une différence négative importante de corrélation est toutefois peu courante et la majorité des cellules présentent une différence positive.



(a) Cellule polaire : hémisphère nord (cellule $n^{\circ}1$)



(b) Cellule équatoriale (cellule $n^{\circ}1050$)



(c) Cellule polaire : hémisphère sud (cellule $n^{\circ}2016$)

FIGURE 3.5 – Différence de corrélation entre cellules (modèle forcé - modèle non forcé)

Les régions équatoriales possèdent, en grande partie, une plus grande corrélation lorsque la structure utilisée provient du modèle forcé. La contrainte de l'interpolation par les données y sera donc d'autant plus importante.

Données standardisées

Les structures de corrélation présentées précédemment sont tirées des données simulées annuelles brutes¹. La matrice de corrélation utilisant des données standardisées présente des valeurs étonnantes entre les pôles (figure 3.6). Le pôle Nord est anticorrélé avec le pôle Sud lors de l'utilisation de ces données (-0.63). Il est toutefois peu probable que cette corrélation soit présente dans le climat réel. La structure utilisée sera donc celle se basant sur les données brutes.

1. Sans aucune modification de normalisation / standardisation

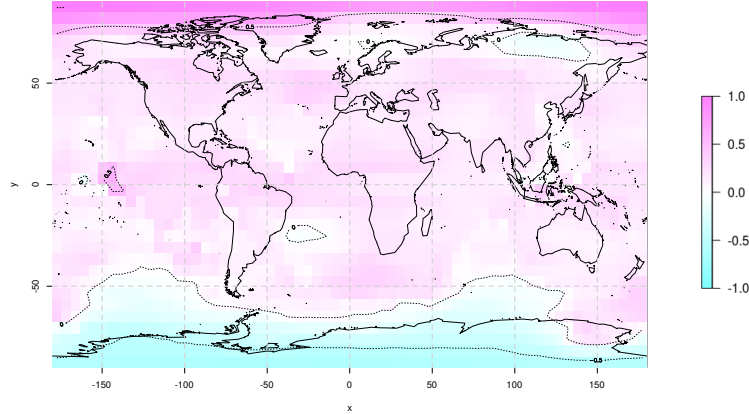


FIGURE 3.6 – Corrélation sur base de données standardisées (cellule $n^{\circ}1$)

3.2.2 Composantes principales

L'analyse en composantes principales des données du modèle *LOVECLIM* (avec ou sans forçages) permet d'évaluer des vecteurs et des valeurs propres qui permettent de résumer la variabilité des données. Les trois premières composantes permettent de regrouper 40% de la variabilité totale, autant pour le modèle forcé que pour le modèle non forcé (39% pour le modèle avec forçages et 38% pour le modèle sans forçages) (table 3.1).

TABLE 3.1 – Valeurs propres *LOVECLIM*

	Modèle forcé	Modèle non forcé
Composante 1	25%	24%
Composante 2	8%	9%
Composante 3	6%	5%

Les cartes des vecteurs propres correspondant aux trois premières composantes sont également très proches, bien que légèrement différentes (figure 3.7). Les valeurs propres les plus élevées des premiers et seconds modes sont localisées proches du pôle. Les troisièmes modes présentent, quant à eux, les zones de valeurs propres élevées au niveau du pôle Sud, à l'ouest de l'Antarctique. Certaines structures sont toutefois uniquement présentes sur le modèle forcé. C'est le cas pour une partie de l'océan Pacifique.

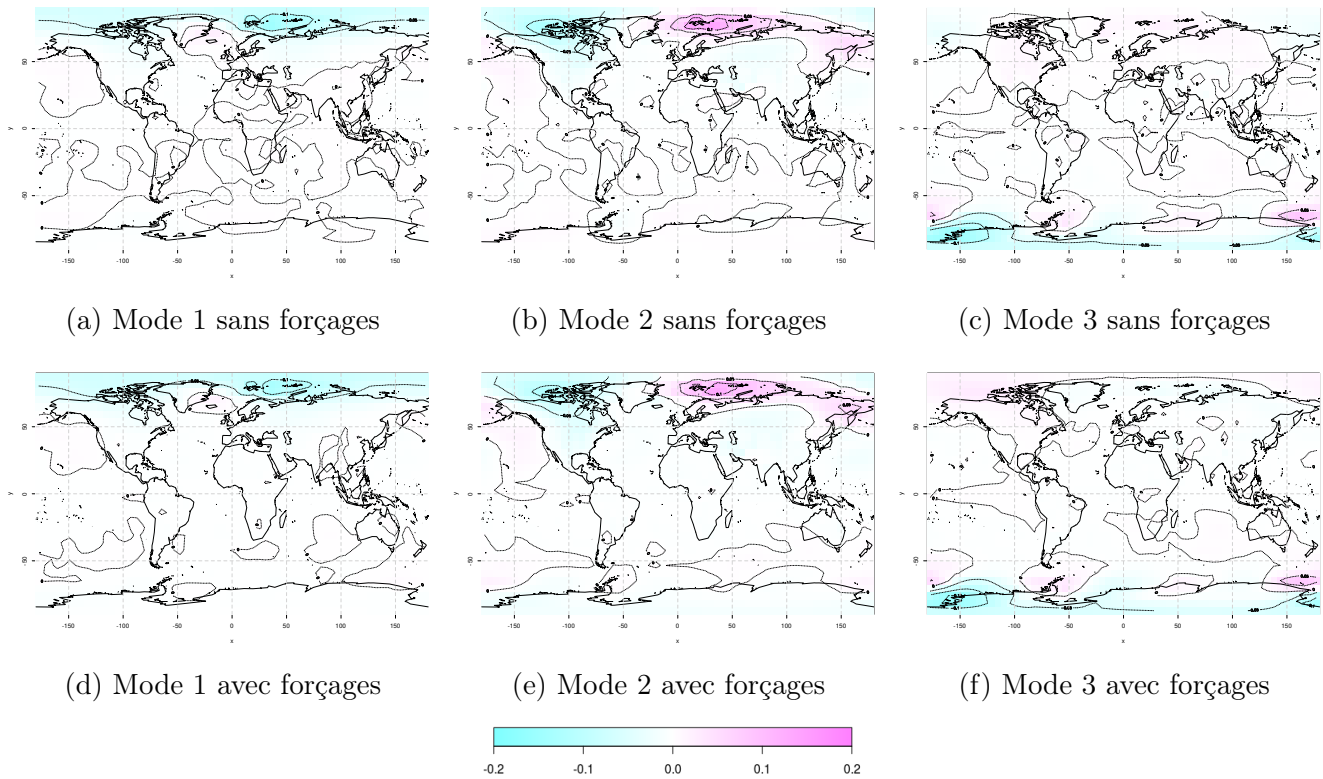


FIGURE 3.7 – Modes propres des modèles

3.2.3 Reconstruction

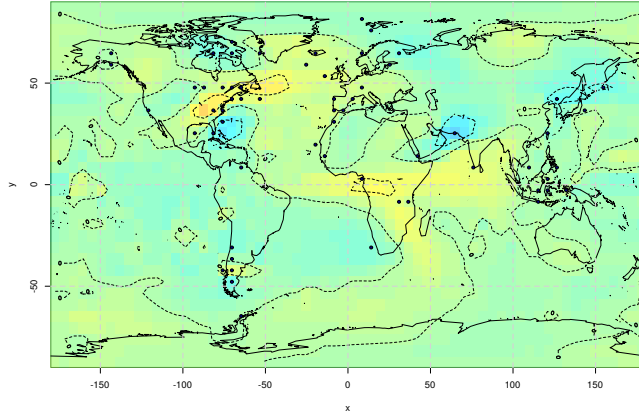
Le second défi de ce travail était de reconstruire un champ spatial de températures grâce aux processus gaussiens.

L'utilisation de la structure de corrélation du modèle *LOVECLIM* permet de disposer d'un rayon d'influence indépendant de la distance pour chacune des observations. Ce rayon d'influence est dépendant de la corrélation entre la cellule d'intérêt et les autres cellules. Il n'est donc plus nécessaire de spécifier une fonction *a priori* définissant le lien entre distance et corrélation.

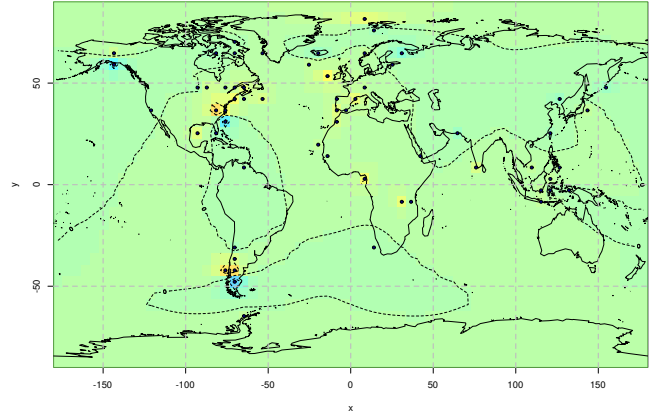
Ce changement de méthode quant à la gestion du lien entre cellules est particulièrement intéressant pour de grandes régions homogènes. Certains points sont représentatifs d'une grande région tandis que d'autres points sont uniquement représentatifs des cellules les plus proches (figure 3.8). Par exemple, le point au sud de l'Inde n'influence quasiment pas les régions situées au nord de sa position, ce qui est notamment dû à la présence de l'Himalaya. Il est toutefois possible d'observer que ce point permet de fournir une information sur une grande partie de l'océan Indien.

Lors de l'utilisation du modèle exponentiel décroissant, les changements des valeurs interpolées sont les plus marqués entre les valeurs de L 500 et 1000. Les différences entre 1000 et 1500 sont, quant à elles, beaucoup plus faibles.

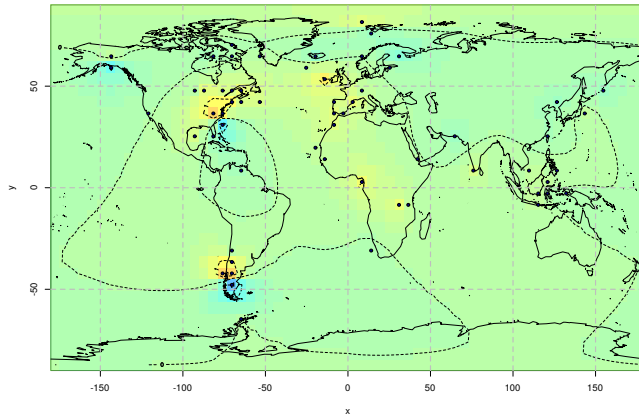
La différence entre les deux méthodes est particulièrement marquée au niveau du golfe du Mexique. Lors de l'utilisation d'un modèle exponentiel décroissant, les archives de cette zone semblent fournir une information sur les régions au sud. Lors de l'utilisation de la structure de *LOVECLIM*, ces archives semblent influencer une partie de l'océan Atlantique proche, notamment vers le nord-est. L'océan Pacifique et la partie sud de l'océan Atlantique sont également relativement uniformes.



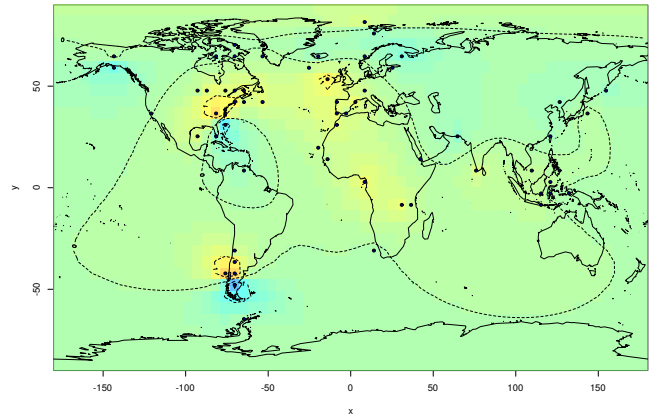
(a) Reconstruction grâce à la structure de *LOVECLIM*



(b) Reconstruction classique ($L = 500$)



(c) Reconstruction classique ($L = 1000$)



(d) Reconstruction classique ($L = 1500$)



FIGURE 3.8 – Exemples de reconstruction pour l'année 501

L'intérêt de l'utilisation des processus gaussiens est de pouvoir exprimer la reconstruction sous une forme probabiliste, définie par une moyenne et une covariance, et dont il est possible d'effectuer des tirages. Cela implique également que la variance de la reconstruction varie en fonction de l'espace et du temps.

Au plus la variance est faible, au plus les tirages se rapprochent de la moyenne. Si cette variance est élevée, alors les histoires sont fort différentes les unes des autres. Lors de l'utilisation de la structure de corrélation obtenue grâce au modèle forcé, cette variance est particulièrement faible à proximité des observations. La répartition de la variance est également homogène sur la durée de l'interpolation. L'Europe, l'est de l'Amérique du Nord, l'équateur et le pôle Nord présentent ainsi des valeurs faibles de variances (figure 3.9). D'autres régions présentent régulièrement une variance élevée, notamment l'Antarctique. Cette variance importante provient du manque de données dans cette région. La seule observation disponible est située sur la péninsule Antarctique, zone très peu corrélée avec l'intérieur du continent.

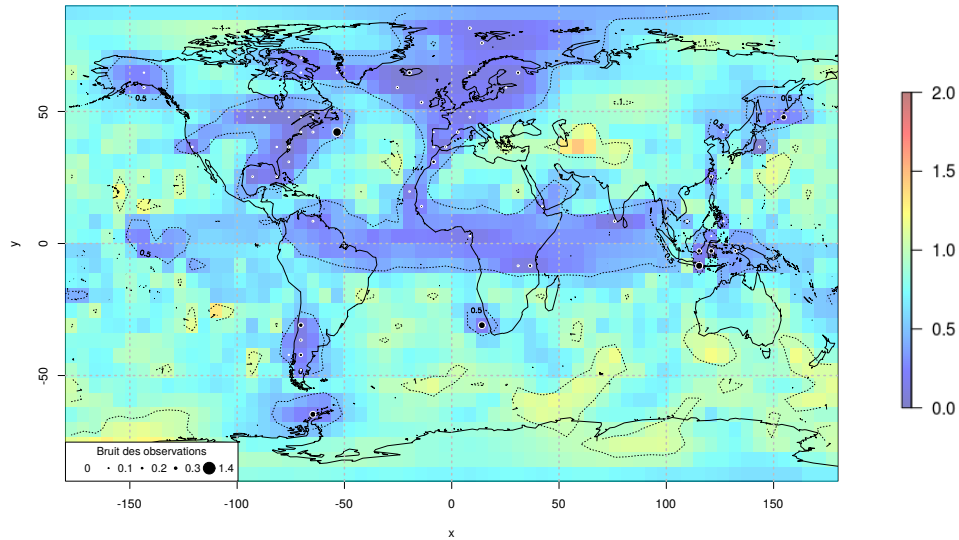


FIGURE 3.9 – Variance de la reconstruction (modèle non paramétrique)

Cette variance peut également être déterminée pour la reconstruction utilisant une paramétrisation de la distance (figure 3.10). La variance y est beaucoup plus élevée. Il est toutefois étonnant que les régions de variance les plus importantes soient localisées proches des points. En effet, il est attendu que cette variance augmente avec l'augmentation de la distance par rapport aux observations. Cette augmentation avec la distance est liée à une diminution de la contrainte appliquée par les observations.

Cette constatation est valable pour différents facteurs d'échelles mais pas pour l'ensemble des points de données. Les points localisés sur le continent européen et sur le continent africain ne présentent pas cette augmentation extrêmement importante de la variance proche des points.

Il est possible que cette variance importante soit liée à la confiance accordée aux observations en Amérique du Nord. Premièrement, la majorité des observations sont contiguës et possèdent un bruit très faible. Le modèle est donc fortement contraint par les données dans cette région. La distance étant faible, la corrélation entre ces cellules est théoriquement importante.

Deuxièmement, les séries peuvent être fort différentes, ce qui pose problème lors de la détermination d'une cellule adjacente aux deux données. Le modèle est donc fortement contraint (bruit faible) lors de l'interpolation par des valeurs différentes mais qui disposent d'un poids important (corrélation importante).

En comparaison avec la méthode utilisant la structure de *LOVECLIM*, la variance est beaucoup plus importante lors de l'utilisation d'une paramétrisation de la corrélation en fonction de la distance. Le modèle exponentiel décroissant simple semble présenter des difficultés à utiliser les données empiriques fournies lors d'une interpolation.

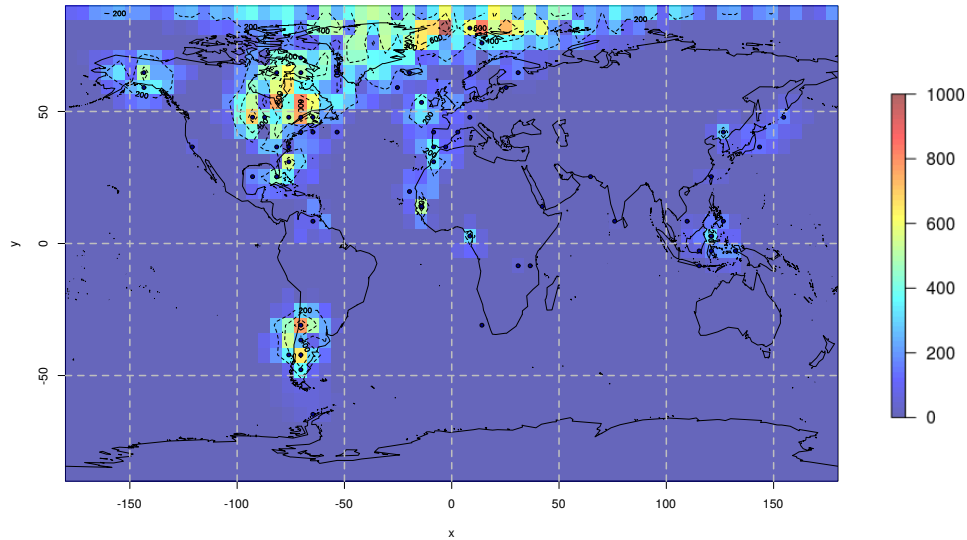
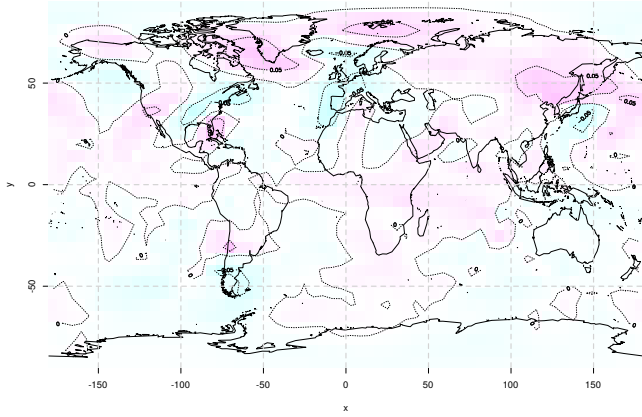


FIGURE 3.10 – Variance de la reconstruction (modèle paramétrique, $L = 5000$)

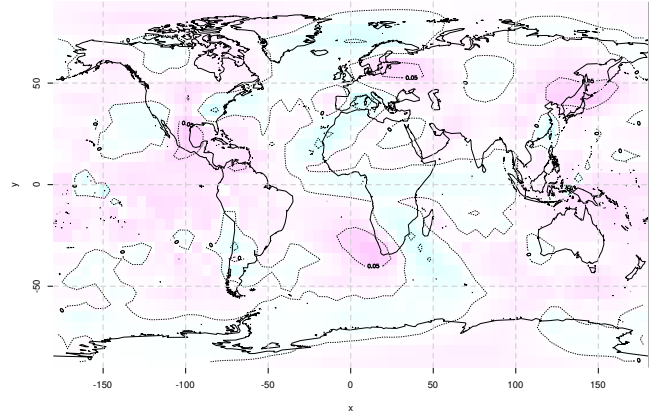
3.2.4 Modes propres de la reconstruction

La contrainte du modèle par les données permet de fournir une reconstruction année par année. Afin de résumer les principales régions de variabilité, une analyse en composante principale est également réalisée. Les trois premiers modes sont très différents des trois premiers modes du modèle seul. La contrainte par les données influence de façon importante la variabilité du modèle, bien que le bruit sur ces dernières soit élevé.

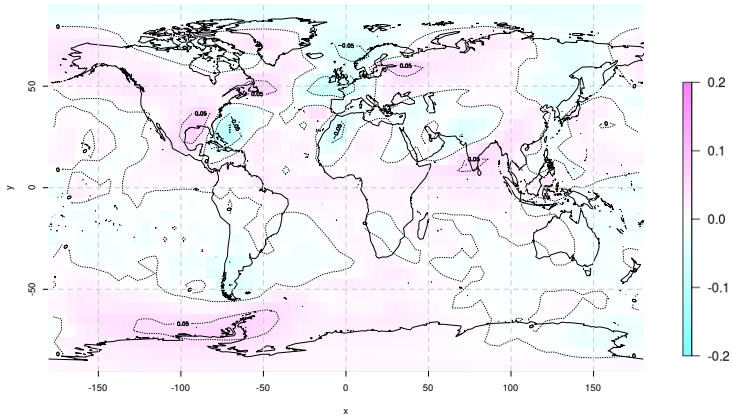
Certaines régions sans données sont reprises dans des modes de variabilités qui n'étaient pas présents dans les modes propres du modèle *LOVECLIM*. La contrainte du *prior* semble donc s'appliquer sur des régions plus ou moins importantes en fonction de la localisation des observations. Cette influence est notamment visible au niveau de la péninsule Antarctique. Le modèle utilisé comme *prior* présente en effet un mode de variabilité relativement circulaire sur cette zone. Après l'application de la contrainte, un mode est toujours présent mais il présente une forme ovale est-ouest.



(a) Mode 1



(b) Mode 2



(c) Mode 3

FIGURE 3.11 – Modes propres de la reconstruction

Chapitre 4

Discussion

4.1 Séries temporelles

4.1.1 Influence du choix d'un modèle exponentiel

Le passage d'un modèle exponentiel décroissant à un modèle exponentiel décroissant élevé au carré influence le remplissage des séries temporelles (équation 4.1 et figure 4.1). Les variations de la courbe moyenne sont beaucoup moins abruptes suite au changement de modèle.

$$C(t_1; t_2) = \exp\left(-\frac{\|t_1; t_2\|^2}{L^2}\right) \quad (4.1)$$

La variance des tirages est également réduite par rapport à la reconstruction faisant appel à une paramétrisation de la distance. L'utilisation d'un modèle exponentiel élevé au carré permet de limiter les changements abrupts de la courbe moyenne. La contrainte causée par les valeurs aberrantes est réduite par rapport à la contrainte appliquée par le modèle exponentiel simple. La valeur de τ relativement faible peut donc être utilisée pour reconstruire des séries temporelles lorsque l'exponentielle est élevée au carré.

Ces variations plus lentes du climat correspondent en effet mieux à l'objectif recherché, à savoir la variabilité basse fréquence. Bien que l'utilisation de ce modèle semble plus réaliste, il n'est pas simple de déterminer si ce type de modèle est plus pertinent que celui proposé par les processus gaussiens classiques avec une optimisation de la longueur d'échelle. Le choix dépendra de l'échelle temporelle étudiée. L'optimisation en fonction de la vraisemblance marginale semble pertinente pour remplir les séries présentant des variations à très long terme. Lorsque le modèle choisi suit une exponentielle décroissante au carré et la valeur moyenne τ tirée de l'autocorrélation des données *LOVECLIM*, le processus reconstruit correspond à une variation à moyen terme.

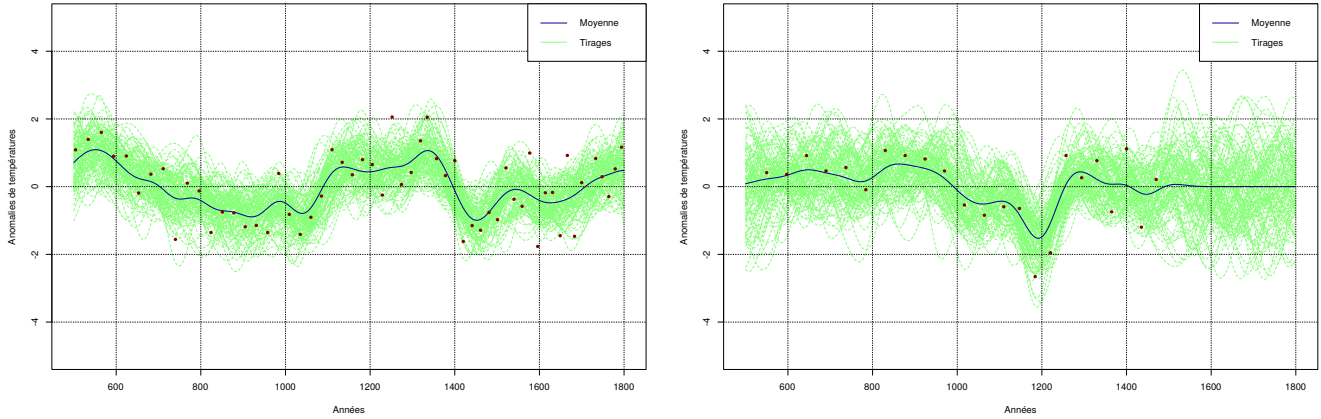


FIGURE 4.1 – Séries temporelles remplies (modèle exponentiel au carré)

Il est toutefois important de noter que lors de l’optimisation de la vraisemblance marginale, les valeurs de L sont dépendantes des séries considérées et non pas de la cellule dans laquelle se trouve la donnée. Cela peut être particulièrement utile si les données utilisées présentent des résolutions temporelles extrêmement différentes (cernes d’arbres et sédiments marins par exemple). Il pourrait alors être plus pertinent de séparer ces deux archives lors du calcul de la longueur d’échelle. La moyenne des valeurs L devra alors se réaliser en fonction du type d’archives. Lors de l’utilisation des données d’un modèle climatique, cette notion de résolution temporelle des données est occultée.

4.1.2 Influence du bruit sur le modèle exponentiel

Les résultats ont présenté l’optimisation de la vraisemblance marginale comme la méthode fournissant un remplissage à très long terme. Le bruit n’est pas un paramètre à optimiser lors de l’utilisation de la vraisemblance marginale, contrairement à ce qu’il est possible de faire lorsque c’est l’autocorrélation qui est utilisée. Ce bruit influence toutefois les résultats de la valeur L (figure 4.2). Le choix de cette valeur peut donc être délicat.

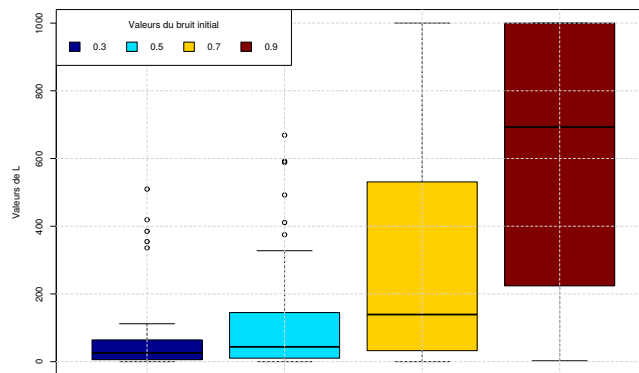


FIGURE 4.2 – Sensibilité au bruit du paramètre L (modèle exponentiel simple)

L’utilisation des données *LOVECLIM* pour déterminer la valeur du paramètre τ peut sembler

discutable.

Toutefois, un paramètre dans le calcul de l'autocorrélation peut permettre de s'ajuster à la résolution temporelle du phénomène étudié. Ce paramètre est la période de lissage des données avant le calcul de l'autocorrélation. En effet, sans lissage, la prévisibilité grâce à l'autocorrélation est assez faible car la variabilité interannuelle est importante. Le choix de la période de lissage est néanmoins relativement libre. Il est pourtant important de noter que la valeur τ est dépendante du choix de la période de lissage (figure 4.3).

En théorie, l'augmentation de la période de lissage permet d'avoir un τ moyen plus élevé et donc de pouvoir s'attarder sur des processus à basses fréquences. Toutefois, l'augmentation de cette période de lissage supprime également une partie de l'information disponible dans les séries temporelles du modèle. Lorsqu'un filtre de 400 ans est utilisé, la quantité d'informations liées à la physique dans une série temporelle est fortement réduite. De plus, une grande période de lissage implique aussi une augmentation de la variance des valeurs τ , notamment causée par une augmentation des valeurs maximales. Les valeurs minimales de τ atteignent rapidement la saturation. Le choix d'une période de lissage revient donc à spécifier une valeur de τ *a priori*.

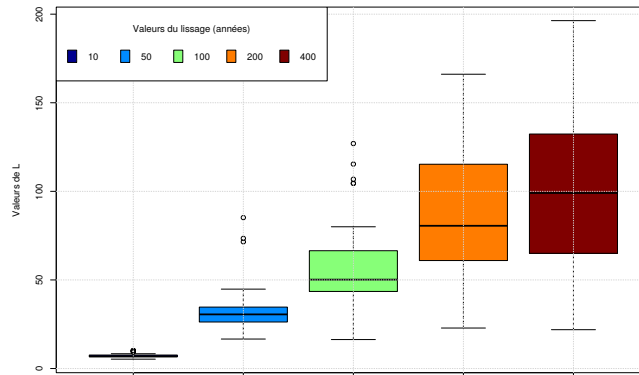


FIGURE 4.3 – Sensibilité au lissage du paramètre L (modèle autocorrélation avec forçages)

4.1.3 Influence du type de *proxy*

La valeur du paramètre de décroissance exponentielle (L ou τ) dépend de la méthode utilisée. Lors de l'utilisation de la vraisemblance marginale, le calcul est directement réalisé sur les données des séries temporelles des archives. La valeur dépend donc des séries mais pourrait également dépendre du type d'archives considérées (sédiments marins ou lacustres). Une légère différence des valeurs L est observable lorsque les deux archives sont traitées séparément. La résolution temporelle des archives, bien que théoriquement différente, ne semble pas influencer le calcul du paramètre d'échelle. Les valeurs moyennes sont de 299 et 311. La grande différence lors du remplissage des séries temporelles est minimale. L'intérêt de séparer ces deux types d'archives lors de leur traitement semble donc limité. Il est également important de préciser que le nombre de sédiments marins (57) est plus important que le nombre de sédiments lacustres (32). Cette différence influence probablement la distribution des valeurs.

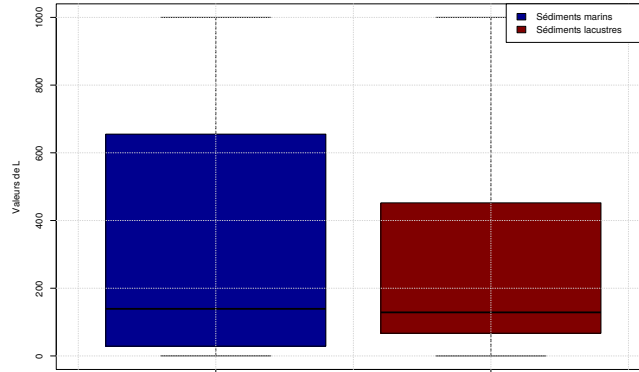


FIGURE 4.4 – Valeurs de L en fonction du type d'archives

4.2 Champ spatial

4.2.1 Influence de la période considérée pour *LOVECLIM*

Modes propres du modèle forcé

L'impact de la longueur des séries temporelles peut s'analyser grâce aux variations des trois premiers modes propres.

Le premier mode est relativement similaire pour les trois périodes différentes. Les valeurs positives et négatives les plus importantes sont localisées dans l'hémisphère nord, proche du pôle. Les différences sont quant à elles présentes hors des hautes latitudes de l'hémisphère nord. Une période courte tend à être plus hétérogène qu'une période longue pour les trois modes.

Le deuxième mode est extrêmement similaire pour les périodes 1-1800 et 1000-1800. Encore une fois, les similarités les plus importantes sont situées dans l'hémisphère nord. Les différences dans l'hémisphère sud sont toutefois peu nombreuses. Le deuxième mode de la période 1700-1800, bien que proche des deuxièmes modes des autres périodes, présente toutefois des valeurs propres plus élevées sur le pôle Nord. La partie nord-est du continent asiatique, liée aux régions scandinaves dans les deux premiers modes des périodes plus longues, est séparée en deux ensembles. Une connexion entre les deux hémisphères est également localisée sur l'océan Pacifique, élément qui n'est pas présent lorsque les périodes de calcul des corrélations sont plus longues.

Le troisième mode est celui qui présente le plus grand nombre de différences parmi les trois périodes. Il est toutefois possible de repérer des similarités entre les deux périodes les plus longues, cette fois dans l'hémisphère sud. Les autres parties du monde, notamment les régions équatoriales sont très différentes.

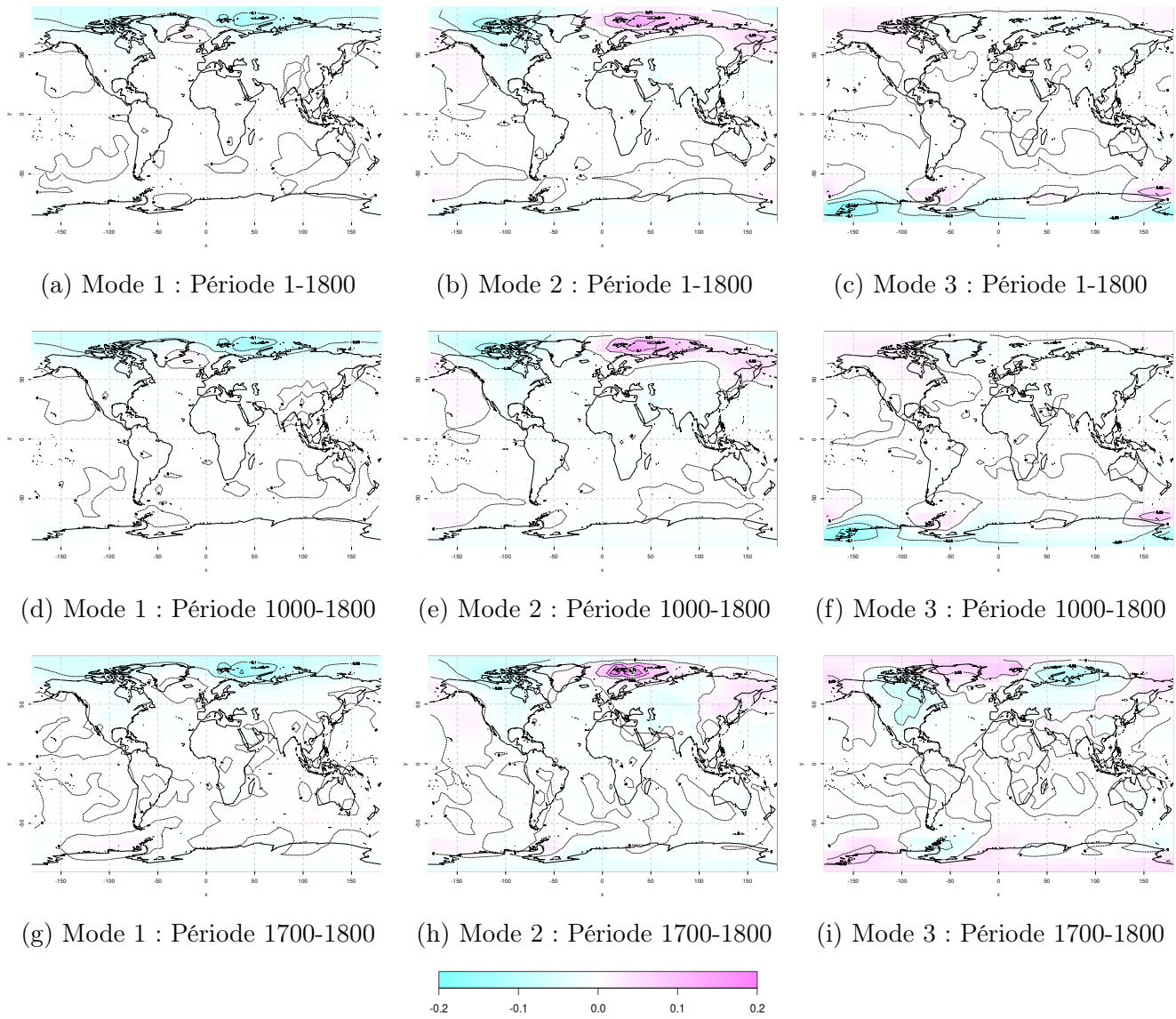


FIGURE 4.5 – Modes propres en fonction de la longueur des données *LOVECLIM*

L'utilisation d'une période plus courte que la période considérée pour la reconstruction semble difficilement justifiable. En effet, la structure de corrélation peut être tronquée par des événements ponctuels indépendants de la variabilité naturelle. L'utilisation d'une structure plus longue peut être pertinente pour inscrire la reconstruction dans un contexte plus large de variation climatique.

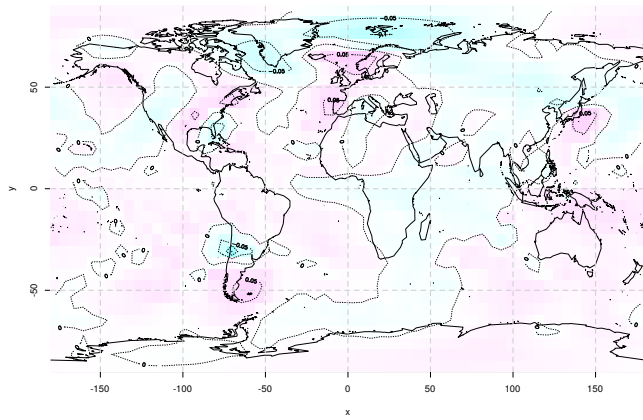
4.2.2 Influence du remplissage sur la reconstruction spatiale

Le choix d'un modèle exponentiel élevé au carré a des conséquences sur le remplissage des séries temporelles. Ce changement influence également la reconstruction du champ spatial. Cette différence est marquée sur les variations de la moyenne au cours des 1300 années mais également sur les trois premiers modes propres (figure 4.6). Cette différence provient des variations plus lisses du remplissage. La contrainte sur la structure de corrélation est donc réduite.

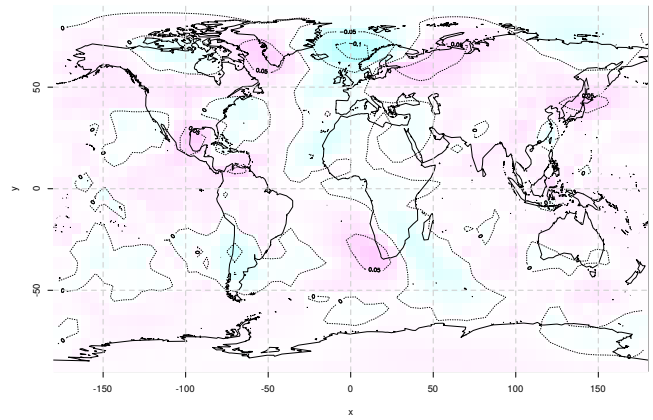
Bien que les premiers modes soient relativement identiques, les différences se marquent sur les

deuxièmes et troisièmes modes. Sur la figure 4.6b, un mode de variabilité important est situé sur le nord de l'Europe. Ce mode n'était pas du tout présent lors de l'utilisation des séries remplies grâce au modèle exponentiel non élevé au carré.

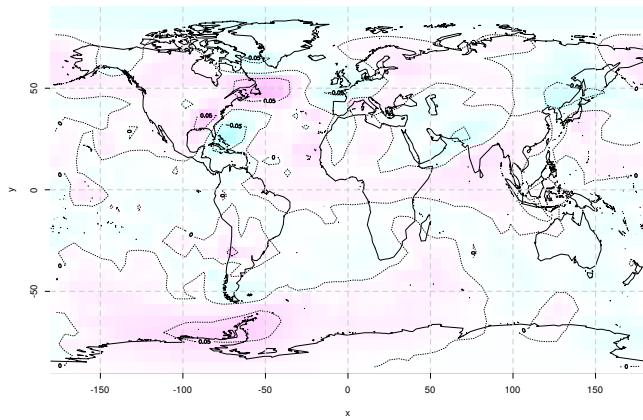
La variance des tirages est également plus faible lorsque les séries utilisées font appel à un modèle exponentiel élevé au carré. Les valeurs extrêmes qui étaient présentes sur les régions polaires de l'hémisphère sud ne sont plus présentes.



(a) Mode 1



(b) Mode 2



(c) Mode 3

FIGURE 4.6 – Modes propres de la reconstruction faisant appel à un remplissage des séries avec un modèle exponentiel au carré

Chapitre 5

Conclusion

Les reconstructions paléoclimatiques sont dépendantes de la qualité des archives utilisées. Les séries temporelles de températures déterminées sur base de ces observations sont généralement incomplètes et bruitées. L'utilisation des équations prédictives des processus gaussiens permettent la réalisation de séries continues intégrant l'incertitude des observations et la covariance entre les années. Cette covariance est paramétrée par une exponentielle décroissante classique. L'utilisation de cette exponentielle implique des changements abrupts des courbes de températures lors du remplissage. Ces changements abrupts sont fortement réduits lorsque le modèle exponentiel est élevé au carré. Cette conclusion est valable pour les deux méthodes de calcul du paramètre d'échelle L , à savoir l'optimisation par vraisemblance marginale et l'utilisation de l'autocorrélation du modèle *LOVECLIM*. La combinaison de l'optimisation par vraisemblance marginale et d'un modèle exponentiel est pertinente pour des variations à basses fréquences. L'utilisation des données du modèle *LOVECLIM* n'est pertinente que lors de leur association à un modèle exponentiel carré. Les séries sont alors remplies par un processus à moyenne fréquence.

Les résultats de l'utilisation des processus gaussiens classiques pour remplir les séries sont dépendants du bruit spécifié *a priori*. Il est en effet complexe de convertir une appréciation personnelle en une valeur précise. Ce problème peut toutefois être supprimé si l'incertitude de chacune des observations était intégrée avec les séries temporelles. Dans notre cas, chaque série temporelle présente le même bruit car le nombre de paramètres libres à optimiser serait trop important pour être calculable (autant de paramètres que de points de données). Toutefois, l'intégration d'un bruit variable devrait rester possible s'il est déterminé en même temps que la conversion entre *proxy* et valeurs de températures.

L'utilisation de ces séries temporelles dans le cadre de la reconstruction d'un champ spatial est également abordée. Une nouvelle méthode d'interpolation permettant la contrainte d'un modèle climatique par des données est développée. Cette méthode permet à l'interpolation de s'affranchir d'une paramétrisation de l'influence d'un point en fonction de la distance. Cette influence est alors déterminée par la corrélation entre les séries temporelles du modèle *LOVECLIM*. L'utilisation de cette structure impose une contrainte plus ou moins importante sur le modèle en fonction de la localisation des données. Cette contrainte est particulièrement importante lors de l'interpolation de la région intertropicale. D'un point de vue plus général, les zones avec la plus grande variance sont situées au niveau de l'hémisphère sud, proches de l'Antarctique. Il est toutefois intéressant d'observer que certaines zones sans observations, comme l'océan Atlantique ou le désert du Sahara ne font pas partie de ces zones à fortes variances.

L'application de cette méthode serait particulièrement intéressante pour des combinaisons à l'échelle continentale. Cette réduction de l'échelle d'étude devrait permettre une meilleure interprétation des valeurs d'anomalies et une possible comparaison avec les variations régionales connues. L'augmentation de la résolution spatiale du modèle utilisé pourrait également améliorer la contrainte du modèle par les données. En effet, un grand nombre d'archives sont regroupées en *clusters*, ce qui empêche l'utilisation d'une partie de l'information car les données sont regroupées dans une même cellule.

Il serait également intéressant d'appliquer ce type de méthode dans d'autres contextes physiques. La structure de corrélation pourrait être fournie par des séries temporelles de données satellites (par exemple de pollution de l'air). Cette structure de corrélation peut ensuite être contrainte par les observations des stations de mesures afin de réaliser une interpolation et éventuellement des prédictions.

Table des figures

1.1	Concentrations de différents gaz déterminées par un carottage lors du programme <i>EPICA</i> [Nieto-Barajas, 2018]	2
1.2	Archives et <i>proxies</i> climatiques [USGS, 2020]	3
1.3	Formation d'un <i>proxy</i> climatique	5
1.4	Fonction depuis une équation $f(x) = 5 - x^2$	7
1.5	<i>Prior</i> dans le cas d'un processus gaussien (moyenne = 0, déviation =1) [Knagg, 2019]	7
1.6	Restriction des fonctions grâce aux données [Knagg, 2019]	8
2.1	Modèle <i>LOVECLIM</i> : Composantes et interactions [Goosse et al., 2010]	10
2.2	Évolution des concentrations en CO_2 et de la température [Goosse et al., 2010]	11
2.3	Localisation des archives climatiques	12
2.4	Autocorrélation <i>LOVECLIM</i> , exponentielle et seuils de significativité	15
2.5	Localisation des points de grille <i>LOVECLIM</i> et cellules les plus proches des données empiriques	16
2.6	Matrice de données pour l'ACP [Bjornsson and Venegas, 1997]	18
3.1	Valeurs de L en fonction du modèle	20
3.2	Série temporelle remplie	21
3.3	Séries temporelles : incohérences	21
3.4	Valeurs du bruit optimisé par vraisemblance marginale (modèle forcé)	22
3.5	Différence de corrélation entre cellules (modèle forcé - modèle non forcé)	23
3.6	Corrélation sur base de données standardisées (cellule $n^{\circ}1$)	24
3.7	Modes propres des modèles	25
3.8	Exemples de reconstruction pour l'année 501	26
3.9	Variance de la reconstruction (modèle non paramétrique)	27
3.10	Variance de la reconstruction (modèle paramétrique, $L = 5000$)	28
3.11	Modes propres de la reconstruction	29
4.1	Séries temporelles remplies (modèle exponentiel au carré)	31
4.2	Sensibilité au bruit du paramètre L (modèle exponentiel simple)	31
4.3	Sensibilité au lissage du paramètre L (modèle autocorrélation avec forçages)	32
4.4	Valeurs de L en fonction du type d'archives	33
4.5	Modes propres en fonction de la longueur des données <i>LOVECLIM</i>	34
4.6	Modes propres de la reconstruction faisant appel à un remplissage des séries avec un modèle exponentiel au carré	35

Bibliographie

- [Attorre et al., 2007] Attorre, F., Alfo', M., Sanctis, M. D., Francesconi, F., and Bruno, F. (2007). Comparison of interpolation methods for mapping climatic and bioclimatic variables at regional scale. *International Journal of Climatology*, 27(13) :1825–1843.
- [Bjornsson and Venegas, 1997] Bjornsson, H. and Venegas, S. (1997). A manual for eof and svd. *McGill University*.
- [Chen et al., 2014] Chen, N., Qian, Z., Nabney, I. T., and Meng, X. (2014). Wind Power Forecasts Using Gaussian Processes and Numerical Weather Prediction. *IEEE Transactions on Power Systems*, 29(2) :656–665.
- [Cook, 1992] Cook, G. C. (1992). Effect of global warming on the distribution of parasitic and other infectious diseases : a review. *Journal of the Royal Society of Medicine*, 85(11) :688–691.
- [Driesschaert, 2005] Driesschaert, E. (2005). Climate change over the next millennia using LOVE-CLIM, a new Earth system model including the polar ice sheets. page 222.
- [Fang and Li, 2016] Fang, M. and Li, X. (2016). Paleoclimate data assimilation : Its motivation, progress and prospects. *Science China Earth Sciences*, 59(9) :1817–1826.
- [Fischer et al., 2004] Fischer, H., Kumke, T., Lohmann, G., Miller, H., and Negendank, J., editors (2004). *The Climate in Historical Times : Towards a Synthesis of Holocene Proxy Data and Climate Models*. GKSS School of Environmental Research. Springer-Verlag, Berlin Heidelberg.
- [Gershman and Blei, 2012] Gershman, S. J. and Blei, D. M. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, 56(1) :1–12.
- [Goosse et al., 2010] Goosse, H., Brovkin, V., Fichefet, T., Haarsma, R., Huybrechts, P., Jongma, J., Mouchet, A., Selten, F., Barriat, P.-Y., Campin, J.-M., Deleersnijder, E., Driesschaert, E., Goelzer, H., Janssens, I., Loutre, M.-F., Morales Maqueda, M. A., Opsteegh, T., Mathieu, P.-P., Munhoven, G., Pettersson, E. J., Renssen, H., Roche, D. M., Schaeffer, M., Tartinville, B., Timmermann, A., and Weber, S. L. (2010). Description of the Earth system model of intermediate complexity LOVECLIM version 1.2. *Geoscientific Model Development*, 3(2) :603–633.
- [Gould, 1967] Gould, P. R. (1967). On the Geographical Interpretation of Eigenvalues. *Transactions of the Institute of British Geographers*, (42) :53–86. Publisher : [Royal Geographical Society (with the Institute of British Geographers), Wiley].
- [Hofert, 2013] Hofert, M. (2013). On Sampling from the Multivariate t Distribution. *The R Journal*, 5(2) :129.
- [IPCC, 2014] IPCC (2014). Climate change 2014 : Synthesis report. contribution of working groups i, ii and iii to the fifth assessment report of the intergovernmental panel on climate change. *Intergovernmental Panel on Climate Change*.
- [Jones and Mann, 2004] Jones, P. D. and Mann, M. E. (2004). Climate over past millennia. *Reviews of Geophysics*, 42(2). _eprint : <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2003RG000143>.

- [Kilbourne et al., 2017] Kilbourne, H., Yu, Z., Neukom, R., Nash, D., Gergis, J., Steig, E. J., Ge, Q., McKay, N. P., Kaufman, D. S., Curran, M. A., Thomas, E. R., Sigl, M., Thirumalai, K., Emile-Geay, J., Chen, M.-T., Seidenkrantz, M.-S., Turney, C., Jacques, J. S., Linderholm, H. W., Horiuchi, K., Björklund, J., Severi, M., Cook, E., Bertler, N., Isaksson, E., eugene wahl, Leduc, G., Martrat, B., Tierney, J. E., Goosse, H., Thamban, M., DeLong, K., Anchukaitis, K., Zinke, J., Uemura, R., Abram, N. J., Shao, X., Dixon, D., von Gunten, L., Wang, J., Addison, J., Evans, M. N., Henley, B., Zhixin, H., McGregor, H. V., Pederson, G. T., Stenni, B., Werner, J., Xu, C., Divine, D., Dixon, B. C., Mundo, I. A., Nakatsuka, T., Phipps, S. J., Routson, C., Tyler, J. J., Allen, K. J., Chase, B., de Jong, R., Ekaykin, A. A., Ersek, V., Filipsson, H. L., Francus, P., Freund, M., Frezzotti, M., Gaire, N., Gajewski, K., Gornostaeva, A., Grosjean, M., Hormes, A., Husum, K., Selvaraj, K., Kawamura, K., Nalan, K., Lorrey, A., Mikhalenko, V., Mortyn, G. P., Motoyama, H., Moy, A., Mulvaney, R., Munz, P., Oerter, H., Opel, T., Orsi, A., Ovchinnikov, D., Porter, T., Roop, H., Saenger, C., Sano, M., Sauchyn, D., Saunders, K., Sicre, M.-A., Sinclair, K., George, S. S., Thapa, U., Viau, A., Vladimirova, D., and White, J. (2017). Pages2k.v2.0.0.rdata. *PAGES2K Database*.
- [Knagg, 2019] Knagg, O. (2019). An intuitive guide to Gaussian processes. <https://towardsdatascience.com/an-intuitive-guide-to-gaussian-processes-ec2f0b45c71d>.
- [Larousse, 2020] Larousse (2020). Dictionnaire de la langue française. *LAROUSSE*. consulté le 06/02/2020 sur [https://www.larousse.fr/dictionnaires/francais/diatom%*c3*%*a9e*/25271?q=diatom%*c3*%*a9e*#25156](https://www.larousse.fr/dictionnaires/francais/diatom%c3%a9e/25271?q=diatom%c3%a9e#25156).
- [Malcolm et al., 2006] Malcolm, J. R., Liu, C., Neilson, R. P., Hansen, L., and Hannah, L. (2006). Global Warming and Extinctions of Endemic Species from Biodiversity Hotspots. *Conservation Biology*, 20(2) :538–548.
- [National Centers for Environmental Information, 2020] National Centers for Environmental Information (2020). State of the climate : Global climate report for annual 2019. *NOAA*. consulté en ligne le 27 janvier sur <https://www.ncdc.noaa.gov/sotc/global/201913>.
- [Nieto-Barajas, 2018] Nieto-Barajas, L. E. (2018). Interpolation of paleoclimatology datasets. *Atmósfera*, 31(2) :125–141.
- [Quiñonero-Candela and Rasmussen, 2005] Quiñonero-Candela, J. and Rasmussen, C. E. (2005). A Unifying View of Sparse Approximate Gaussian Process Regression. *Journal of Machine Learning Research*, 6(Dec) :1939–1959.
- [Rasmussen and Williams, 2006] Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass. OCLC : ocm61285753.
- [Schulz et al., 2018] Schulz, E., Speekenbrink, M., and Krause, A. (2018). A tutorial on Gaussian process regression : Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, 85 :1–16.
- [Thejll and Schmith, 2005] Thejll, P. and Schmith, T. (2005). Limitations on regression analysis due to serially correlated residuals : Application to climate reconstruction from proxies. *Journal of Geophysical Research : Atmospheres*, 110(D18). _eprint : <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2005JD005895>.
- [USGS, 2020] USGS (2020). Paleoclimate : Home. 2020-02-06 sur <https://www2.usgs.gov/landresources/lcs/paleoclimate/>.

- [Widmann et al., 2010] Widmann, M., Goosse, H., van der Schrier, G., Schnur, R., and Barkmeijer, J. (2010). Using data assimilation to study extratropical northern hemisphere climate over the last millennium. *Climate of the Past*, 6(5).
- [Wratt and Mulan, 2007] Wratt, D. and Mulan, B. (2007). Natural variations in climate. 2020-01-29 sur <https://niwa.co.nz/our-science/climate/information-and-resources/clivar/variatiions>.
- [Şen, 2009] Şen, Z. (2009). Global warming threat on water resources and environment : a review. *Environmental Geology*, 57(2) :321–329.

CORENTIN VISÉE

🏠 05/11/1997

@ corvis@hotmail.be

☎ +32 479 19 20 27

📍 Bruyère des Onze Villes, 2, 7050 Jurbise

DIPLÔMES

Master en sciences géographiques, orientation climatologie,
à finalité approfondie

UCLouvain – École de Géographie, Faculté des Sciences

📅 Septembre 2018 – Août 2020 📍 Louvain-la-Neuve, Belgique

Titre du mémoire:

*Interpolation géospatiale des observations de températures entre 500
et 1800AD en tenant compte de simulations climatiques*

Baccalauréat en sciences géographiques

UNamur – Département de Géographie, Faculté des Sciences

📅 Septembre 2015 – Juin 2018 📍 Namur, Belgique

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Faculté des sciences

Place des sciences, 2 bte L6.06.01, 1348 Louvain-la-Neuve, Belgique | www.uclouvain.be/sc