MASTER 2 MATHÉMATIQUES APPLIQUÉES STATISTIQUES

RAPPORT DE STAGE :

Mélange de placettes temporaires et permanentes pour le suivi de la biodiversité : approche statistique par simulation

Auteur :
FAYE EL HADJI CISSE

Superviseurs :
Frédéric GOSSELIN, INRAE Nogent-Sur-Vernisson
Romain JULLIARD, CESCO – MNHN
Fabien LAROCHE, INRAE Toulouse
Antoine LEVEQUE, UMS PatriNat
Didier CHAUVEAU, Institut Denis Poisson - Universite d'Orleans

03/2022 - 09/2022

Résumé

De nombreux programmes de suivis utilisent des données de comptage pour quantifier la tendance temporelle pour pouvoir statuer sur l'état de la population et mettre en place des plans de conservation et de gestion. Ces suivis nécessitent des plans d'échantillonnages. L'objectif principal de ce présent travail est de proposer un modèle statistique complet pour l'estimation de la tendance temporelle d'une population en partant des travaux réalisés par Rhodes & Jonzen, 2011. Les champs aléatoires gaussiens du modèle statistique sont approximés en utilisant des champs aléatoires gaussiens de Markov grâce à l'approche des équations différentielles partielles stochastiques (SPDE). Ce travail s'articule autour de 2 sections. La première vise à comparer des plans d'échantillonnage avec uniquement de placettes permanentes, uniquement de placettes temporaires et un mélange de placettes permanentes et temporaires pour l'estimation de la tendance temporelle. La meilleure stratégie d'échantillonnage est fonction de l'autocorrélation temporelle et du paramètre d'échelle lié à l'autocorrélation spatiale et diffère selon la présence ou l'absence d'effet de nuggets. La seconde section de ce présent travail concerne la comparaison de modèles statistiques pour l'estimation de la tendance temporelle. Les performances relatives des méthodes sont évaluées en termes d'erreur de type I et de précision réelle des estimations.

Mots-clés :
autocorrélation temporelle, autocorrélation spatiale, effet de nuggets, équations aux dérivées partielles stochastiques, plan d'échantillonnage, placettes, tendance temporelle.

Remerciements

Ce travail est le fruit de la combinaison d'efforts de plusieurs personnes. Je remercie le tout puissant qui, par sa grâce nous a permis d'arriver au bout de nos efforts en nous donnant la santé, la force, le courage et en nous faisant entourer des merveilleuses personnes dont je tenais à remercier.

Je remercie :

Romain Julliard, Antoine Leveque, Frédéric Gosselin et Fabien Laroche, mes maîtres de stage, de m'avoir proposé un sujet d'étude riche et ambitieux, et de m'avoir accordé leur confiance tout au long du stage. Je leur exprime toute ma gratitude pour leur encadrement sans faille, leur rigueur au travail, leurs multiples conseils, leurs orientations et leur disponibilité malgré leurs multiples occupations ; Un grand merci à Frédéric et Fabien pour les vives discussions très constructives, que nous avons eues ensemble ;

Didier Chauveau, mon tuteur académique, pour ses remarques et ses suggestions très constructives ;

Tous les enseignants de la filière Master Mathématiques Appliquées, Statistiques, de l'université d'Orléans pour leurs multiples conseils et leurs efforts déployés afin de nous assurer une formation de qualité ;

Tout le personnel du Muséum National d'Histoire Naturelle de m'avoir accueilli, dans une ambiance amicale et détendue ;

A l'équipe BiosP, particulièrement Thomas Opitz, Denis Allard et Julien Papaix pour nos échanges très intéressent portant sur le sujet du présent travail ;

Nouédéhouénou Houessou pour nos échanges et ses conseils concernant la rédaction du présent rapport ;

Ma famille et mes amis pour leurs encouragements et leurs prières durant toute la formation ;

Mes collègues du master MAS de l'université d'Orléans ;

Tous ceux qui de près ou de loin ont contribué à l'accomplissement de ce travail.

Table des matières

Liste des tableaux

Table des figures

Chapitre 1

Introduction générale

1.1 MNHN, CESCO et Projet PASSIFOR-2

J'ai réalisé, du mois de Mars au mois de Septembre 2022, mon stage de fin d'études au sein du MNHN (Muséum d'Histoire Naturelle Nationale, sous la direction de Romain Julliard, Antoine Leveque, Frédéric Gosselin et Fabien Laroche . J'ai intégré l'équipe du Centre d'Ecologie et des Sciences de la Conservation (CESCO) qui développe des recherches sur la conservation de la biodiversité à travers des approches multidisciplinaires en écologie des populations et en sciences sociales. Pour comprendre les mécanismes à l'origine du déclin de la biodiversité, il est nécessaire de décrire ses variations au cours du temps et dans l'espace. Pour effectuer cette description, un travail important réside dans l'élaboration d'indicateurs de suivi de la biodiversité. Le calcul de ces indicateurs doit permettre de mesurer, le plus simplement et efficacement possible, la réponse de la biodiversité à différentes menaces ou actions humaines. Mettre au point des indicateurs de suivi de la biodiversité permet ensuite de réaliser des études prospectives et d'élaborer des scénarios de son évolution dans différentes conditions environnementales. C'est une aide à la décision dans les choix politiques, collectifs ou individuels.

Mon stage s'est éffectué dans le cadre du projet PASSIFOR-2 (Propositions pour l'Amélioration du Système de Suivi de la biodiversité FORestière). C'est un projet de recherche appliquée et d'expertise sur des maquettes de suivi de la biodiversité forestière. Il vise à une identification, un état des lieux et une analyse des réseaux de suivi forestier et/ou de biodiversité. L'objet de ce projet est d'élaborer différentes « maquettes » (assemblages d'éléments existants et à créer) de suivi de la biodiversité en forêt. Il vise une aide aux politiques publiques dans le domaine du suivi de la biodiversité, centré sur la forêt en lien avec les autres milieux. La surveillance de la biodiversité fait référence à un système d'observations régulières de l'écosystème au fil du temps, informant sur l'état de la biodiversité dans le but de détecter et d'évaluer les tendances de quantités telles que la richesse en espèces, la diversité des espèces, l'abondance des espèces (Mitusova 2006, Gosselin et al. 2007).

1.2 Problématique statistique

Face au déclin global de la biodiversité, des suivis de populations animales et végétales sont réalisés sur de grandes zones géographiques et durant une longue période afin de comprendre les facteurs déterminant la distribution, l'abondance et les tendances des populations. Ces suivis à larges échelles permettent de statuer quantitativement sur l'état des populations et de mettre en place des plans de gestion appropriés de la biodiversité.

Une population peut se définir comme un ensemble d'individus d'une même espèce occupant une zone géographique commune et se reproduisant entre eux (voir Millstein 2010). Elle se caractérise par plusieurs paramètres tels que son effectif, son taux de croissance (augmentation, stabilité, diminution de l'effectif), sa densité (le nombre d'individus par unité de surface) et sa répartition (aire occupée par la population). Ces paramètres sont des éléments clés pour déterminer son état et en particulier sa viabilité (Shaffer 1981). Suivre une population à large échelle, comme par exemple à l'échelle d'un pays ou d'un continent, impose des contraintes non négligeables et nécessite potentiellement des moyens logistiques et /ou financiers importants (voir Dickinson et al. 2010 ; Jones 2011). Dans la grande majorité des cas, il sera impossible de suivre l'ensemble de la zone et il faudra donc échantillonner seulement un sous-ensemble représentatif de celle-ci.

La plupart des modèles suppose explicitement que l'abondance d'une population peut être bien décrits comme constant dans l'espace. Il est de plus en plus reconnu que l'hétérogénéité spatiale a des impacts sur la dynamique des populations. Thorson et al. [18] ont montré que les modèles non spatiaux peuvent entraîner des estimations biaisées de la densité-dépendance si l'abondance est autocorrelée dans l'espace. En revanche, le modèle spatial fournit des estimations précises de la densité-dépendance. Thorson et al. [18] ont adapté le modèle de Gompertz pour approximer les abondances locales de la population sur un espace continu à l'aide de champs aléatoires gaussiens.

Les travaux de Rhodes & Jonzen [16] ont montré que les stratégies optimales d'échantillonnage pour estimer la tendance temporelle d'une population dépendaient de manière cruciale des niveaux d'autocorrélation spatiale et temporelle : si la tendance temporelle est constante sur un pas de temps important (ici : 30 ans), quand l'autocorrélation spatiale est faible et la temporelle forte, il est préférable d'échantillonner beaucoup de sites peu souvent visités et inversement si l'autocorrélation spatiale est forte et la temporelle faible. Toutefois, Rhodes & Jonzen n'ont pas considéré les placettes temporaires et leur travail était basé sur de nombreuses simplifications et hypothèses. Parmi ces limitations/hypothèses, citons :

- La tendance de l'espèce est constante dans le temps sur un long pas de temps (30 ans), cette tendance pouvant néanmoins varier dans l'espace de manière aléatoire et non structurée.
- Les niveaux d'autocorrélation temporelle et spatiale sont connus. Or dans un vrai suivi, ces paramètres sont à estimer via une approche statistique complète estimant tous les paramètres inconnus.

Partant des travaux de Rhodes & Jonzen [16], Houessou (2021) [15] a identifié les limites et les hypothèses simplificatrices du modèle de Rhodes et Jonzen et proposé les extensions nécessaires pour une application dans des écosystèmes réels. Houessou a affermi la base mathématique du travail de Rhodes & Jonzen et mis à jour les hypothèses et limitations de leur travail. Son travail s'articule autour du mélange de placettes temporaires asynchrones avec des placettes permanentes synchrones. Il a trouvé que la proportion de placettes temporaires optimale pour la d'estimation de la tendance temporelle était assez forte pour des niveaux d'autocorrélation spatiale moyen-forts et des niveaux d'autocorrélation temporelle moyens. Malgré tout, l'amélioration avec l'introduction des placettes non permanentes de la précision de l'estimateur de la tendance temporelle était très ténue (baisse de 1 à 3% de l'erreur type).

Les résultats issus des travaux de Rhodes & Jonzen et Houessou ne sont pas établis sur une approche statistique car une part importante des paramètres est considérée comme connue. De plus, ils considèrent que la tendance temporelle de l'espèce est constante dans l'espace et dans

le temps. Ils cherchent des valeurs optimales des paramètres de l'échantillonnage (nombre de passages sur les placettes permanentes, proportion de placettes temporaires) parmi un petit nombre de valeurs (optimisation discrète).

L'objectif du travail présent est de proposer un cadre statistique pour l'estimation de la tendance temporelle d'une population. Le modèle spatio-temporel utilisé est le modèle de Gompertz avec des champs aléatoires gaussiens pour modéliser les variations spatiales de l'abondance à l'équilibre, de la tendance temporelle et des variations environnementales. Pour tenir compte des variations indépendantes dans l'espace, nous avons introduit dans nos modèles les effets de nuggets. Dans un premier temps, il s'agira de comparer l'incertitude obtenue pour l'estimation de la tendance temporelle avec uniquement des placettes permanentes, uniquement des placettes temporaires et un mélange équilibré de placettes permanentes et temporaires. Ainsi, nous pouvons vérifier un des résultats du travail de Houessou (2021) [15] dans un cadre statistique. Dans le projet PASSIFOR-2, la consultation de la communauté écologique a fait ressortir la questions des placettes permanentes avec un point de vue globale en faveur des placettes permanentes dans les suivi forestier. Nous nous intéressons également, suivant les valeurs de l'autocorrélation temporelle et du paramètre d'échelle lié à l'autocorrélation spatiale, la meilleure stratégie d'échantillonnage pour l'estimation de la tendance temporelle. Dans la littérature, beaucoup d'écologistes considèrent que la tendance temporelle est constante dans l'espace. Dans un second temps, nous comparons des modèles statistiques dans les cas suivants : (1) la tendance temporelle est constante dans l'espace, (2) elle varie aléatoirement de manière indépendante dans l'espace, (3) la tendance temporelle varie spatialement avec une matrice de covariance de Matèrn.

Chapitre 2

Définitions de certains concepts

Ce premier chapitre est consacré à la définition de certains concepts tels que l'échantillonnage en écologie, l'autocorrélation spatiale, l'autocorrélation temporelle, les placettes et les plans de revisite. On évoque ensuite les données spatiales.

2.1 Pourquoi échantillonner ?

En écologie, il est généralement impossible de mesurer une ou des caractéristiques sur l'ensemble des unités ou de la surface occupée par d'un groupe d'intérêt. Ceci peut résulter de plusieurs causes, telles des contraintes de temps, d'argent ou un manque de personnel qualifié. Ou encore, il peut être impossible de mettre la main sur l'ensemble des individus d'une population. Les variables ou caractéristiques d'intérêt ne sont pas observées ou mesurées sur chacun des individus de la population mais sur un fragment appelé échantillon. Choisir le plan d'échantillonnage consiste à choisir de quelle manière les données seront recueillies sur le terrain. Selon le but visé et les contraintes rencontrées, plusieurs plans d'échantillonnage sont disponibles et répondent à des besoins particuliers. Il existe deux grandes catégories de techniques (méthodes) d'échantillonnage, les méthodes d'échantillonnage non probabilistes et les méthodes d'échantillonnage probabilistes : (voir Amel Meddad-Hamza [1] et Levy & al. [12]).

- L'échantillonnage probabiliste fait référence à la sélection d'un échantillon d'une population lorsque cette sélection repose sur le principe de la randomisation, c'est-à-dire la sélection au hasard ou aléatoire.
- L'échantillonnage non probabiliste est une méthode qui consiste à sélectionner des unités dans une population en utilisant une méthode non aléatoire ou non probabiliste.

Dans ce travail, on utilise un échantillonnage probabiliste avec des mailles d'échantillonnage comme dans l'inventaire forestier national (IFN). Pour cela, on transpose sur la zone d'étude une grille de cellules régulières de même taille. Le nombre de cellules dépend de la taille de l'échantillon et peut être supérieur à la taille de l'échantillon. Puis, on tire aléatoirement le nombre de cellules faisant partie de l'échantillonnage. Ensuite dans chaque cellule, on choisit l'emplacement des sites de manière aléatoire. Ce type d'échantillonnage permet d'éviter les phénomènes de concentration des placettes à certains endroits de l'espace d'étude et permet de faire varier la distance inter-placettes afin de pouvoir mieux estimer l'autocorrélation spatiale à faible portée.

2.2 Autocorrélations spatiale et temporelle

L'autocorrélation spatiale peut être définie comme la ressemblance des valeurs prises par une variable aléatoire, exprimée en fonction de leur localisation géographique. Elle est omniprésente dans un large éventail de données écologiques. L'autocorrélation spatiale positive signifie que "les données qui sont proches les unes des autres dans l'espace sont souvent plus similaires que celles qui sont éloignés" (Cressie, 1993). L'analyse de l'autocorrélation permet d'estimer la portée de la dépendance spatiale afin de le prendre en compte dans l'analyse des données. La corrélation temporelle quant à elle reflète la ressemblance entre les informations recueillies à deux dates différentes sur le même site.

2.3 Placettes et plan de revisite

En écologie, une placette est un point ou site d'observation permettant de recueillir un nombre important de données. Chaque placette est caractérisé par ces coordonnées (longitude et latitude) qui permettent de la localiser sur le domaine d'étude. Dans ce travail, on définit une *placette permanente* comme une placette échantillonnée au moins 2 fois durant toute l'étude. Si elle est échantillonnée une seule fois durant toute l'étude, on parlera de *placette temporaire*.

Il est important de définir un plan de revisite lors d'un suivi. Les placettes temporaires sont échantillonnées de manière aléatoire et asynchrone alors que les placettes permanentes sont échantillonnées de manière répété, à des intervalles réguliers ou variables. Durant le suivi, les placettes permanentes peuvent être synchroniques ou asynchrones. Notons qu'il existe différents plans revisites (voir McDonald [14] pour plus de détails).

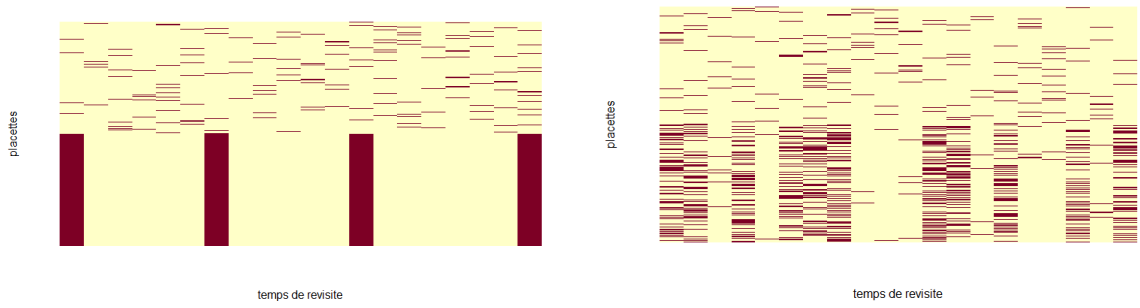

FIGURE 2.1 – Mélange de placettes : à gauche placettes permanentes synchrones et à droite placettes permanentes asynchrones

2.4 Données Spatiales

Les données spatio-temporelles sont définies comme la réalisation d'un processus stochastique indexé par l'espace et le temps :

$$X(s,t) = \{x(s,t), (s,t) \in \mathcal{D} \times \mathbb{N}\} \tag{2.1}$$

où $\mathcal{D}$ est un sous-ensemble (fixe) de $\mathbb{R}^2$. Selon que $\mathcal{D}$ soit une surface continue ou une collection dénombrable d'unités spatiales à d dimensions, le problème peut être défini comme un processus aléatoire spatial continu ou discret, respectivement.

Si nous supposons que $X(s,t)$ est continu dans l'espace et a des lois marginales gaussiennes, nous avons un champ gaussien continuement indexé (GF). Pour compléter la spécification de la distribution de $x_t(s)$ ($x_t(s) = x(s,t)$), il est nécessaire de définir sa covariance spatio-temporelle :

$$Cov(x_t(s_i), x_q(s_j)) = \sigma^2 M(s_i, s_j|t, q) \tag{2.2}$$

où σ^2 est la variance et $M(s_i, s_j|t, q)$ est la fonction de corrélation spatio-temporelle ($M(s_i, s_i|t, t) = 1$ pout tout s_i et t).

En fonction des hypothèses, la fonction de covariance spatio-temporelle peut être adaptée à chaque situation. En cas de la stationnarité dans l'espace et dans le temps, la fonction de covariance spatio-temporelle peut être spécifiée comme une fonction de la distance spatiale Δ_{ij} et du décalage temporel $\delta_{tq} = |t - q|$, et elle est donc définie comme suit

$$Cov(x_t(s_i), x_q(s_j)) = \sigma^2 M(\Delta_{ij}; \delta_{tq}) \tag{2.3}$$

Si nous supposons la séparabilité, la fonction de covariance spatio-temporelle est donnée par

$$Cov(x_t(s_i), x_q(s_j)) = \sigma^2 M_1(\Delta_{ij}) M_2(\delta_{tq}) \tag{2.4}$$

avec M_1 et M_2 étant les fonctions de corrélation spatiale et temporelle, respectivement. Alternativement, il est possible de considérer une fonction de covariance purement spatiale donnée par

$$Cov(x_t(s_i), x_t(s_j)) = \sigma^2 M(\Delta_{ij}) \tag{2.5}$$

Nous devons définir la fonction de covariance spatiale de Matèrn qui contrôle la corrélation spatiale à la distance $||h|| = ||s_i - s_j||$, et cette covariance est donnée par :

$$M(h|\nu, \kappa) = \frac{2^{1-\nu}}{\Gamma(\nu)} (\kappa||h||)^\nu K_\nu(\kappa||h||) \tag{2.6}$$

où K_ν est la fonction de Bessel modifiée de second espèce et d'ordre $\nu > 0$. Le paramètre ν mesure le degré de lissage du processus et est généralement maintenu fixe ($\nu = 1, 2, ..$). Dans la suite, on pose $\nu = 1$. A l'inverse, $\kappa > 0$ est un paramètre d'échelle spatiale liée à la portée r, c'est à dire la distance à laquelle la corrélation spatiale devient presque nulle. Généralement, la portée désigne la distance à laquelle la corrélation spatiale est proche de 0.1. Dans le cas de corrélation de Matèrn, pour tout κ, cette portée est donnée par $r = \dfrac{\sqrt{8\nu}}{\kappa}$.

Chapitre 3

Matériel et Méthodes

On introduit dans ce chapitre la modélisation spatio-temporelle de la dynamique de population par un modèle state-space de Gompertz. On s'intéresse ensuite à l'utilisation des progrès récents dans le domaine des champs aléatoires gaussiens pour l'estimation des GRF de nos modèles. On se penchera sur les différents scénarios de simulations envisagés. Dans la suite de ce présent travail, on se met dans un cadre purement simulatoire.

3.1 Modèle state-space

Les modèles state-space sont une modélisation populaire d'analyse des données de séries chronologiques en écologie. Ils sont une approche flexible pour modéliser la dynamique des populations. Ce sont des modèles de type hiérachique et leur structure permet la modélisation de deux processus stochastiques : un processus latent et un processus observé. Ces processus sont reliés par des relations de dépendances probabilistes :

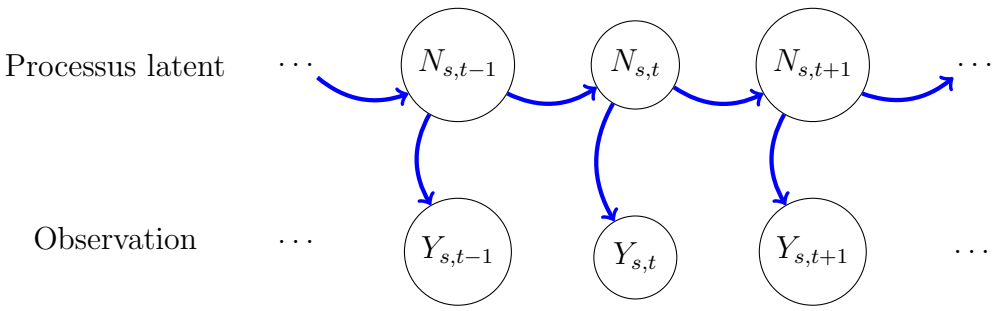

FIGURE 3.1 – Structure de dépendance dans un modèle state-space

— Le processus latent n'est pas observé et tente de refléter l'état réel mais caché du phénomène étudié. Il contient le processus dynamique de la population.
— Le processus observé est une série chronologique de mesures liées au processus latent. Les observations sont conditionnelles au processus latent.

Ces deux composants stochastiques agissent à différents niveaux de la hiérachie du modèle, et le cadre state-space permet de les modéliser séparément. (voir Auger-Méthé et al. [3] pour plus d'informations)

3.1.1 Processus latent

Le modèle de Gompertz est souvent utilisé pour modéliser la dynamique des populations dépendant de la densité en écologie et représente bien la dynamique d'un large gamme d'espèces.

Dans ce travail, la dynamique de la population a été représentée par le modèle de Gompertz stochastique en temps discret :

$$N_{s,t} = N_{s,t-1} exp\Bigg(-0.5\sigma_u^2 + \rho(\log(K_{s,t-1}) - \log(N_{s,t-1})) + u_{s,t} \Bigg) \tag{3.1}$$

où $N_{s,t}$ est l'abondance de la population à l'instant t au niveau du site s et $K_{s,t}$ l'abondance à l'équilibre (ou capacité de charge) de la population à l'instant t au niveau du site s. Le paramètre ρ définit le niveau de la densité dépendance. $u_{s,t}$ représente la variation environnementale stochastique du taux de croissance de la population au site s à l'instant t et le terme σ_u^2 représente sa variance.

A l'échelle logarithmique, le modèle de Gompertz se présente comme un modèle autorégressif d'ordre 1 ($AR(1)$) :

$$\log(N_{s,t}) = \log(N_{s,t-1}) - 0.5\sigma_u^2 + \rho\Big(\log(K_{s,t-1}) - \log(N_{s,t-1}) \Big) + u_{s,t} \tag{3.2}$$

Comme Rhodes et Jonzen ([16]), nous avons supposé qu'à l'échelle logarithmique, il avait une tendance temporelle linéaire déterministe de l'abondance à l'équilibre de la population due à un processus tel que la perte de l'habitat ou le changement climatique. Cela se traduit par :

$$\log(K_{s,t}) = \log(K_{s,t-1}) + R_s = \log(K_s^0) + tR_s \tag{3.3}$$

où R_s est la tendance temporelle et K_s^0 la capacité de charge initiale au niveau du site s .

Cela aboutit au modèle :

$$\log(N_{s,t}) = \log(N_{s,t-1}) + \rho\Big(\log(K_s^0) + (t-1)R_s - \log(N_{s,t-1}) \Big) - 0.5\sigma_u^2 + u_{s,t} \tag{3.4}$$

Il est également nécessaire spécifier une condition initiale, représentant l'abondance de la population lorsque $t = 0$. Dans ce travail, à $t = 0$, nous avons :

$$\log(N_{s,1}) = \log(K_s^0) + \phi \tag{3.5}$$

où ϕ est le log-rapport de l'abondance à t = 0 et de la médiane de la distribution de l'abondance à l'équilibre.

Pour des raisons de simplifications, nous posons $\mathbf{N}_t = (N_{1,t}, .., N_{S,t})$, $\mathbf{K}^0 = (K_1^0, .., K_S^0)$, $\mathbf{R} = (R_1, .., R_S)$ et $\mathbf{u}_t = (u_{1,t}, .., u_{S,t})$ les vecteurs contenant respectivement l'abondance de la population, l'abondance à l'équilibre, la tendance temporelle et la variation environnementale stochastique du taux de croissance de la population pour les S sites à l'instant t.

On peut donc écrire

$$log(\mathbf{N}_t) = \begin{cases} \log(\mathbf{K}^0) + \phi\mathbf{1} & \text{si } t = 0 \\ \log(\mathbf{N}_{t-1}) + \rho\Big(\log(\mathbf{K}^0) + (t-1)\mathbf{R} - \log(\mathbf{N}_{t-1}) \Big) - 0.5\sigma_u^2\mathbf{1} + \mathbf{u}_t & \text{si } t > 0 \end{cases} \tag{3.6}$$

Nous avons supposé que $\log(\mathbf{K}^0)$, $\mathbf{R}$ et $\mathbf{u}_t$ sont des champs aléatoires gaussiens avec une fonction de covariance spatiale $M(.|\nu, \kappa)$ identique définie en (2.6) :

$$\log(\mathbf{K}^0) \sim MVN\Big(m_{\log(\mathbf{K}^0)}, \Sigma_{\log(\mathbf{K}^0)} \Big) \textbf{ avec } \Sigma_{\log(\mathbf{K}^0)} = \sigma^2_{\log(\mathbf{K}^0)} M(h|\nu, \kappa)$$

$$\mathbf{R} \sim MVN\Big(m_{\mathbf{R}}, \Sigma_{\mathbf{R}} \Big) \textbf{ avec } \Sigma_{\mathbf{R}} = \sigma^2_{\mathbf{R}} M(h|\nu, \kappa)$$

$$\mathbf{u}_t \sim MVN\Big(0, \Sigma_{\mathbf{u}} \Big) \textbf{ avec } \Sigma_{\mathbf{u}} = \sigma^2_{\mathbf{u}} M(h|\nu, \kappa)$$

Modèle avec grille latente

Lorsque le nombre de placettes est important, il est a priori difficile de modéliser en spatio-temporel sur un temps long. Pour des problèmes numériques, il est important, voire nécessaire de réduire la dimension du problème. Nous avons réfléchi sur un modèle de type state-space avec une dynamique portée par une grille fixe latente, servant à paramétrer les observations sur les placettes (temporaires ou permanentes).

L'idée est de subdiviser l'espace en un maillage carré dont les noeuds du maillage sont indépendants des sites observés. Au lieu de calculer directement la log-abondance sur les sites observés, on calcule cette quantité sur les noeuds de la grille latente puis on effectue une interpolation pondérée pour avoir les log-abondances sur les sites observés.

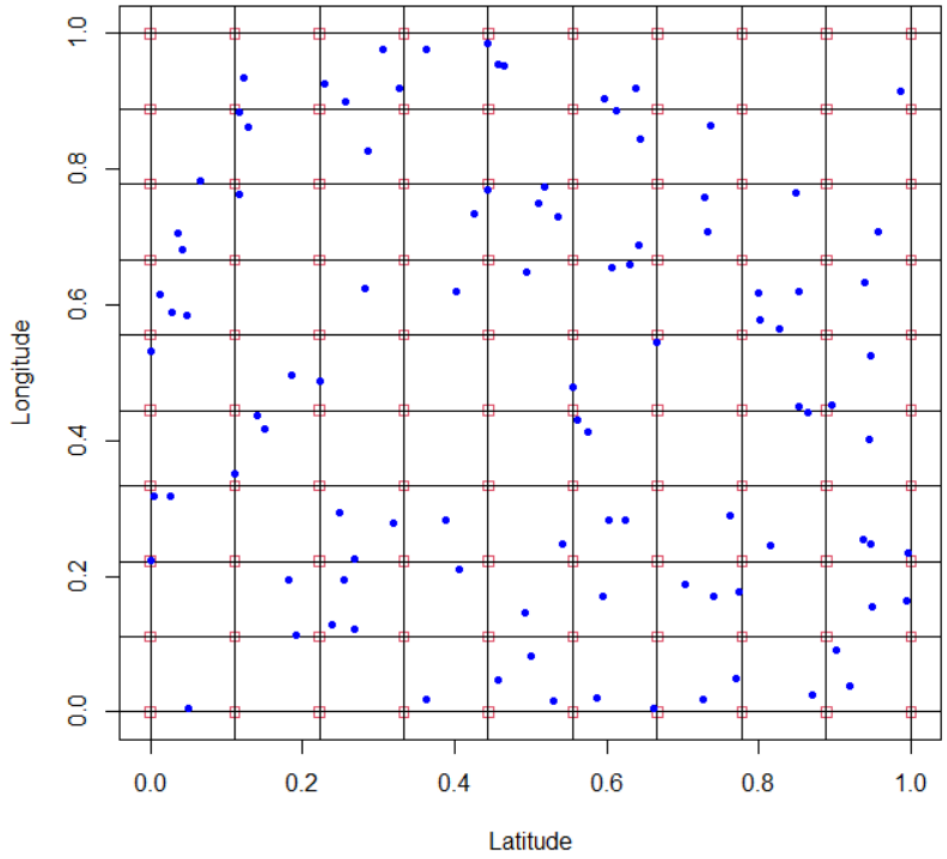

FIGURE 3.2 – Grille latente avec les sites observés en bleu et les noeuds de la grille latente en rouge

Soit r le site observé où l'on veut calculer la log-abondance. Pour cela, on détermine les quatre sites A_r, B_r, C_r et D_r de la grille latente qui entourent le site observé r. Puis, on détermine les nouvelles coordonnées $(x_r = latitude, y_r = longitude)$ du site r dans le nouveau repère orthonormé d'origine A_r ($||A_rB_r|| = 1$ et $||A_rD_r|| = 1$). Les coefficients respectifs de A_r, B_r, C_r

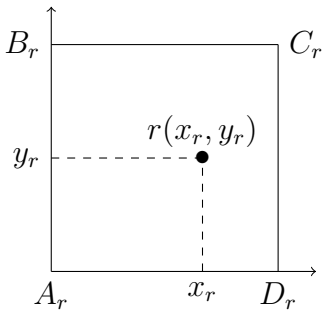

FIGURE 3.3 – Grille latente avec site observé r et les quatre points de la grille latente qui l'entourent

et D_r dans la pondération sont donnés par les formules suivantes :

$$\begin{aligned} coef_{A_r} &= (1 - x_r)(1 - y_r) \\ coef_{B_r} &= (1 - x_r)y_r \\ coef_{C_r} &= x_r y_r \\ coef_{D_r} &= x_r(1 - y_r) \end{aligned}$$

Ensuite, à l'échelle logarithmique, l'abondance de la population au niveau du site observé r à l'instant t est donnée par

$$\begin{aligned} \log(N_{r,t}^{obs}) = coef_{A_r} \log(N_{A_r,t}^{grid}) + coef_{B_r} \log(N_{B_r,t}^{grid}) \\ + coef_{C_r} \log(N_{C_r,t}^{grid}) + coef_{D_r} \log(N_{D_r,t}^{grid}) + \mu_r^0 + (t-1)\mu_r^1 \end{aligned} \tag{3.7}$$

où $\mu_r^0 \sim N(0, \sigma_{\mu^0})$ et $\mu_r^1 \sim N(0, \sigma_{\mu^1})$.

Il existe une différence entre les placettes. Cette différence est due à des variations spatiales indépendantes propre à chaque placette. La différence entre les placettes et les variations spatiales indépendantes représentent les effet de nuggets. Dans le modèle statistique, μ^0 représente l'effet de nuggets mis sur l'abondance à l'équilibre et μ^1 celui mis sur la tendance temporelle.

Sur le plan statistique, pour que le modèle avec la grille latente fonctionne bien, il faut que le pas de la grille latente soit lié à la portée (distance à laquelle la corrélation est de 5%) par la formule $pas \leq \frac{portée}{14}$ (Desassis et al. [7]). Lorsque la portée est faible, le pas de la grille latente ne pourra pas être abaissé pour des raisons numériques et le modèle pourrait ne pas bien fonctionner. Deux solutions sont possibles pour remédier à cela :

- une première solution est de fixer la taille maximale de la grille latente 18×18. A partir de la taille maximale de la grille latente, on déduit la valeur minimale pour la portée. Comme la portée est liée au paramètre d'échelle spatiale κ, nous considérons dans la suite que $\kappa \in \{1.2, 3.5\}$
- Une autre alternative est d'ajouter à chaque noeud de la grille latente principale, de pas $pas.princ$, une sous-grille de pas $pas.sec$ ($pas.sec = \dfrac{pas.princ}{6}$ ou $pas.sec = \dfrac{pas.princ}{7}$ par exemple) de sorte à avoir une matrice de variance-covariance locale petite (de taille 6×6 ou 7×7 par exemple). Toutes les matrices de variance-covariance seraient de même taille et a priori avec le même kappa que la grille latente principale mais avec des variances différentes. Lors de l'interpolation pour les points observés, on ferait la somme de l'interpolation sur la grille principale et sur la grille secondaire. Après l'estimation

des paramètres du modèle, on vérifierait si le *pas.sec* est inférieure à la *portée*/14. Si ce n'est pas le cas, on rajoute une troisième couche de grille avec le même principe noté ci-dessus et ainsi de suite.

Dans ce présent travail, la première solution est utilisée. Le paramètre d'échelle est donnée par la formule suivante :

$$\kappa = \frac{\sqrt{8}}{14c}(n.grid - 1) \tag{3.8}$$

où c est le coté de la grille latente et $n.grid$ le nombre de noeuds de la grille latente. Dans la suite, on utilise un maillage carré de côté 1 comme grille latente, i.e $c = 1$.

3.1.2 Modèle d'observation

En écologie comme dans bien d'autres domaines, les données de comptage sont l'un des types de données le plus couramment utilisées. Cependant, les échantillons de données de comptage comprennent souvent de nombreux zéros et quelques abondances fortes. Souvent ces types de données sont modélisées par un processus de poisson. Dans une distribution de poisson, la variance est égale à la moyenne. Cependant, pour diverses raisons, la quantité de variation pour chaque unité d'échantillonnage est généralement plus élevée que prévu par un processus de poisson. Cette variation supplémentaire, appelée surdispersion, est causée par l'hétérogénéité spatio-temporelle du processus qui produit les données, généralement en raison d'erreurs d'observation ou d'erreurs de processus. Il existe plusieurs alternatives pour modéliser les processus de poisson surdispersés. Une des approches les plus courantes utilisées pour modéliser des données de comptage surdispersées est la distribution négative binomiale. (voir Lindén et Mantyniemi [13])

Soit la variable $Y_{s,t}$ représentant le processus observé de l'espèce au niveau du site s au temps t. Pour tenir compte de la dispersion dans nos données, nous supposons alors que la variable d'observation suit une loi négative binomiale à deux paramètres :

$$Y_{s,t} \sim NegBin(N_{s,t}^{obs}, \sigma^2_{Y_{s,t}}). \tag{3.9}$$

où $N_{s,t}^{obs}$ est l'abondance locale au niveau du site observé s à l'instant t modélisée dans le processus latent. La variance $\sigma^2_{Y_{s,t}}$ s'écrit comme une fonction quadratique de la moyenne $N_{s,t}^{obs}$.

$$\sigma^2_{Y_{s,t}} = (\theta_1 + 1)N_{s,t}^{obs} + \theta_2(N_{s,t}^{obs})^2 \tag{3.10}$$

où θ_1 est le paramètre de croissance linéaire et θ_2 celui de croissance quadratique de la dispersion $\sigma^2_{Y_{s,t}}$.

3.2 Estimation du modèle

Pour l'estimation des paramètres du modèle, nous utilisons les progrès récents dans le domaine des champs gaussiens aléatoires. En effet, les données spatiales ou spatio-temporelles peuvent être traitées avec l'approche des équations différentielles partielles stochastiques (SPDE) proposée par Lindgren et al. (2011) [10]. Il s'agit de représenter nos champs aléatoires gaussiens (GRF) avec une fonction de covariance de Matèrn définie en (2.6) comme des champs aléatoires de Markov gaussiens (GMRF). Cela produit d'importants avantages de calculs. En effet, les GMRF se caractérisent par des matrices de précision creuses, ce qui permet de mettre en oeuvre des méthodes numériques efficaces sur le plan du calcul. Pour un GRMF dans R^2, le coût de calcul est de $\mathcal{O}(n^{3/2})$, ce qui est une amélioration significative par rapport à $\mathcal{O}(n^3)$ des GRF.

3.2.1 Approche des équations différentielles partielles stochastiques (SPDE)

L'approche SPDE permet à un champ gaussien avec la fonction de covariance de Matèrn d'être considéré comme un processus aléatoire spatial discrètement indexé, ce qui présente des avantages considérables en termes de calcul (voir par exemple Lindgren et al. (2011) [10]). Les champs aléatoires gaussiens sont affectés par le *"big n problem"*. Ceci est dû aux coûts de calcul de $\mathcal{O}(n^3)$ lorsqu'il s'agit d'effectuer une opération d'algèbre matricielle avec matrices de covariance $n \times n$ denses, ce coût est nettement plus grand lorsque les données augmentent en espace et en temps. Pour résoudre ce problème, nous utilisons une approximation qui relie un champ gaussien à indexation continue avec des fonctions de covariance de Matèrn à un processus aléatoire spatial à indexation discrète, c'est-à-dire un champ aléatoire gaussien de Markov (GMRF).

L'idée est de construire une représentation finie d'un champ de Matèrn en utilisant une combinaison linéaire de fonctions de base dénommées dans une triangulation d'un domaine $\mathcal{D}$ donné. Le GRF est considérée comme la solution stationnaire d'une SPDE. Ce lien permet de remplacer la fonction de covariance spatiale ou spatio-temporelle et la matrice de covariance dense d'un GRF par une structure de voisinage et une matrice de précision creuse, respectivement, qui sont toutes deux des éléments typiques d'un GMRF. Ceci, à son tour, apporte des avantages considérables en termes de calcul (Lindgren et al. (2011) [10]).

En utilisant l'expression décrite dans (2.6), lorsque $\nu + \dfrac{d}{2}$ est un entier, une représentation linéaire par morceaux, efficace du point de vue du calcul peut être construite en utilisant une représentation différente du champ de Matern $x_t(s)$, à savoir la solution stationnaire du SPDE

$$(\kappa^2 - \Delta)^{\alpha/2} x_t(s) = W(s) \tag{3.11}$$

où d est la dimension spatiale, $\alpha = \nu + \dfrac{d}{2}$ un entier, $\Delta = \sum_{i=1}^{d} \dfrac{\partial^2}{\partial s_i^2}$ est l'opérateur Laplacien et $W(s)$ est le bruit blanc spatial. La variance marginale est liée à la SPDE par

$$\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{d/2}\kappa^{2\nu}\tau^2} \tag{3.12}$$

Une solution approximative du SPDE peut être trouvée en utilisant la méthode des éléments finis. Cette méthode divise le domaine spatial $\mathcal{D}$ en un ensemble de triangles non sécants conduisant à un maillage triangulaire à n nœuds et n fonctions de base. Les fonctions de base $\psi_k()$ sont définies comme des fonctions linéaires par morceaux sur chaque triangle qui est égal à 1 au sommet k et égal à 0 aux autres sommets. Ensuite, le champ gaussien indexé continûment x_t est représenté comme un champ aléatoire de Markov gaussien discrètement indexé (GMRF) au moyen des fonctions de base finies définies sur le maillage triangulaire

$$x_t(s) = \sum_{k=1}^{n} \psi_k(s) x_k, \tag{3.13}$$

où n est le nombre de vertex dans la triangulation, $\{\psi_k\}$ est l'ensemble des fonctions de base et $\{x_k\}$ sont des poids normalement distribués.

3.2.2 Utilisation de R-INLA et TMB

L'approximation des GRF dans nos modèles par des GMRF donne

$$\begin{aligned}\log(\mathbf{K}^0) &\sim MVN(m_{\log(\mathbf{K}^0)}, Q^{-1}_{\log(\mathbf{K}^0)})\\ \mathbf{R} &\sim MVN(m_{\mathbf{R}}, Q^{-1}_{\mathbf{R}})\\ \mathbf{u}_t &\sim MVN(0, Q^{-1}_{\mathbf{u}})\end{aligned}$$

où les matrices de précision sont données par

$$\begin{aligned}Q_{\log(\mathbf{K}^0)} &= \tau^2_{\log(\mathbf{K}^0)}(\kappa^4 C + 2\kappa^2 G_1 + G_2)\\ Q_{\mathbf{R}} &= \tau^2_{\mathbf{R}}(\kappa^4 C + 2\kappa^2 G_1 + G_2)\\ Q_{\mathbf{u}} &= \tau^2_{\mathbf{u}}(\kappa^4 C + 2\kappa^2 G_1 + G_2)\end{aligned}$$

$\kappa > 0$ est un paramètre d'échelle à estimer défini en (2.6) avec $\nu = 1$. La variance marginale de chaque champs aléatoire gaussien est donnée par (3.12).

L'approche SPDE est déja implémenté dans le package R-INLA. Les matrices creuses C, G_1 et G_2 sont approximées par la méthode INLA-SPDE (Integrated Nested Laplace Approximation with Stochastic Partial Differential Equation).

Nous estimons ensuite tous les paramètres à l'aide de Template Model Builder (TMB). TMB est relativement connu de la communauté des statistiques spatiales, car il s'agit d'un outil de modélisation d'effets aléatoires très flexible qui permet aux utilisateurs de définir des modèles d'effets aléatoires complexes via des modèles C++ simples. Il utilise l'approximation de Laplace pour calculer la vraisemblance marginale des paramètres à effet fixe. Les champs gaussiens $\log(\mathbf{K}^0), \mathbf{R}$ et $\mathbf{u}_t$ ont été considérés comme des effets aléatoires.

3.3 Scénarios de simulations

Dans les scénarios décrits ci-dessous, le suivi se fait sur une période de 20 ans durant laquelle le nombre total d'observation B est fixé à 10000. Ce qui correspond à un effort d'échantillonnage uniforme dans le temps avec 500 placettes par an. On fixe le nombre de repassage sur les placettes permanentes à 4 pour avoir une nette différence entre les placettes permanentes et temporaires que dans le cas où le nombre de repassage sur les placettes permanentes est fixé à 2. En effet, Houessou (2021) ([15]) avait fixé le nombre de repassage sur les placettes permanentes à 2 et avais obtenu un gain de précision faible. Dans les simulations, nous considérons que les placettes permanentes sont asynchrones. Et le modèle utilisé est le modèle avec grille latente. L'ensemble des scénarios proposés sont utiles au projet PASSIFOR2, notamment à l'inventaire forestier qui ne repose que sur des placettes temporaires.

3.3.1 Scénario 1

On suppose sur la base des résultats préliminaires de Houessou (2021) ([15]) que l'introduction de placettes temporaires améliore la précision de l'estimateur de la tendance temporelle. Nous souhaitons pouvoir le vérifier dans un contexte statistique complet.

Les paramètres à faire varier de manière continue sont le paramètre d'échelle κ lié à l'autocorrélation spatiale et l'autocorrélation temporelle ρ. Vu que l'autocorrélation temporelle ρ est dans l'intervalle $[0, 1]$ dans Rhodes & Jonzen, nous supposerons que $\rho \sim U(0, 1)$. La proportion de placettes temporaires peut prendre uniquement les valeurs $0, 0.5$ et 1, donc $p \in \{0, 0.5, 1\}$. Le paramètre d'échelle *kappa* est lié au pas de la grille latente. Pour des raisons numériques, κ est dans l'intervalle $[1, 5.5]$: $\kappa \sim U(1, 5.5)$. $\kappa = 1.2$ correspond à une grille latente de taille 7×7 alors que $\kappa = 3.5$ celle d'une grille de taille 18×18 (cf. 3.1.1). Les autres paramètres du modèle restent fixes : $\phi = 3$, $m_R = 0.04$, $m_{\log(K^0)} = 1$, $\sigma_u = 0.4$, $\sigma_R = 0.02$ et $\sigma_{\log(K^0)} = 1$. Les simulations ont été faites sans les effets nugget.

3.3.2 Scénario 2

Ce scénario de simulation concerne l'introduction des effets de nugget dans le modèle statistique et la partie simulation des données : l'hypothèse est que, s'il y a des nuggets, le cas p=1 devrait ne pas bien estimer au moins la part de nuggets avec peut-être des répercussions sur l'estimation du reste des paramètres. Si tel est le cas, c'est important à savoir pour inciter des cas comme l'Inventaire Forestier à introduire des placettes permanentes.

Comme dans le scénario 1, on s'intéresse à la variation de l'erreur type de la tendance temporelle en fonction de ρ et κ pour différentes valeurs de σ_{μ^0} et σ_{μ^1}.

On reprend les même valeurs des paramètres dans le cas du scénario 1 pour cet scénario. Comme ρ et κ, les variances $\sigma_{\mu^0}, \sigma_{\mu^1}$ des nuggets varient de manière continue : $\sigma_{\mu^0} \sim U(0.01, 0.2)$ et $\sigma_{\mu^1} \sim U(0.001, 0.005)$.

On tire au sort 1000 valeurs du couple (ρ,κ) pour le scénario 1 et 1000 valeurs du quadruplet $(\rho, \kappa, \sigma_{\mu^0}), \sigma_{\mu^1})$ puis on génère trois jeux de données d'observation correspondant au cas où $p \in \{0, 0.5, 1\}$.

3.3.3 Scénario 3

Rappelons que le modèle statistique général considéré est :

$$log(\mathbf{N}_t) = \begin{cases} \log(\mathbf{K}^0) + \phi\mathbf{1} & \text{si } t = 0 \\ \log(\mathbf{N}_{t-1}) + \rho\Big(\log(\mathbf{K}^0) + (t-1)\mathbf{R} - \log(\mathbf{N}_{t-1})\Big) - 0.5\sigma^2\mathbf{1} + \mathbf{u}_t & \text{si } t > 0 \end{cases} \tag{3.14}$$

Dans la littérature, les modèles proposés pour l'estimation de la tendance temporelle supposent que cette quantité est constante dans l'espace. Dans ce scénario, on compare la qualité de l'estimation de la tendance temporelle des trois modèles statistiques suivants :

— *modèle 1* : La tendance temporelle est considérée comme constante dans l'espace :

$$\mathbf{R} = m_R\mathbf{1}.$$

— *modèle 2* : La tendance temporelle varie dans l'espace de manière aléatoire et non structurée. On ne prend pas en compte la corrélation entre les placettes pour la modélisation de la tendance temporelle :

$$\mathbf{R} \sim MVN\Big(m_{\mathbf{R}}, \Sigma_{\mathbf{R}}\Big) \textbf{ avec } \Sigma_{\mathbf{R}} = \sigma^2_{\mathbf{R}} I$$

— *modèle 3* : La tendance temporelle est modélisée par un champs gaussien aléatoire avec une matrice de Matèrn. Ainsi, elle varie de manière aléatoire et structurée dans l'espace. Et on prend en compte la corrélation entre les placettes :

$$\mathbf{R} \sim MVN\Big(m_{\mathbf{R}}, \Sigma_{\mathbf{R}}\Big) \textbf{ avec } \Sigma_{\mathbf{R}} = \sigma^2_{\mathbf{R}} M(h|\nu, \kappa)$$

Nous fixons tous les paramètres des modèles : $\rho = 0.5$, $p = 0.5$, $\kappa = 2.5$, $\phi = 3$, $m_R = 0.04$, $m_{\log(K^0)} = 1$, $\sigma_u = 0.4$, $\sigma_R = 0.035$ et $\sigma_{\log(K^0)} = 1$. On ne considère pas les effets de nuggets dans les parties simulation et l'analyse statistique des données. Pour 1000 réplications, on s'intéresse à l'estimation de la tendance temporelle moyenne $\hat{m}_R$, à l'incertitude de l'estimation $SE(\hat{m}_R)$ mais aussi aux erreurs de type I obtenus avec les différents modèles. Dans les sorties graphiques, on va avoir la boite à moustache des estimations de la tendance temporelle moyenne $\hat{m}_R$ pour les différents modèles dont la vraie valeur m_R est représentée en pointillée mais aussi la boite à moustache de l'incertitude de l'estimation $SE(\hat{m}_R)$. On représente aussi graphiquement l'erreur de type I associés aux modèles. De plus les même de graphiques seront représentés pour les paramètres ρ et κ.

3.4 Méthodes d'analyses

3.4.1 Analyse de l'estimation de la tendance temporelle et de l'incertitude

Pour les deux premiers scénarios, on s'intéresse dans un premier temps à l'estimation de la tendance temporelle moyenne $\hat{m}_R$ et de l'incertitude de l'estimation $SE(\hat{m}_R)$ obtenues avec des échantillonnages comportant uniquement des placettes permanentes ($p = 0$), uniquement de placettes temporaires ($p = 1$) et un mélange équilibré de placettes permanentes et temporaires ($p = 0.5$). Le $SE(\hat{m}_R$ est l'approximation de l'erreur type d'estimation de la tendance temporelle , fondée sur les propriétés asymptotiques de l'estimateur du vraisemblance et est calculé numériquement par TMB.

On calcule également les ratios $\Delta^q_{SE} = \dfrac{SE(\hat{m}_R)_q}{SE(\hat{m}_R)_{0.5}}$ avec $q \in \{0, 1\}$ qui seront comparés à 1. Pour chaque valeur de q, on s'intéresse à la variation de la quantité Δ^q_{SE} en fonction de l'autocorrélation temporelle ρ et du paramètre d'échelle κ lié à l'autocorrélation spatiale. Une représentation graphique de la variation de Δ^q_{SE} peut être obtenue avec l'utilisation d'un modèle de régression additive généralisé GAM qui effectue un lissage. Ce lissage rend visible les différences dans la répartition de Δ^q_{SE}. Dans la modélisation du GAM, on prend en compte l'interaction entre ρ et κ. Les résumés des résultats des modèles GAM sont mis en annexe.

Le choix du meilleur plan d'échantillonnage pour l'estimation de la tendance temporelle m_R dans la suite est basé sur les incertitudes d'estimation $SE(m_R)$. Lorsqu'il n'existe pas de différence significative en termes d'erreur standard entre les plans d'échantillonnage, on se penchera sur la variation de la quantité Δ^q_{SE}, $q = 0, 1$, en fonction de ρ et κ pour faire ce choix. On compare de ce fait l'incertitude de l'estimation de la tendance temporelle avec les échantillonnages composés uniquement des placettes permanentes ($q = p = 0$) et uniquement des placettes temporaires ($q = p = 1$) avec un échantillonnage composé de mélange équilibré de placettes permanentes et temporaires suivant les valeurs de ρ et κ.

3.4.2 Erreur de type I

On s'intéresse aux erreurs de type I des estimateurs et plus particulièrement à celle de la tendance temporelle. L'erreur de type I est une mesure de la qualité globale de l'inférence en

termes d'adéquation fréquentiste de l'estimation.

Par exemple, pour le calcul de l'erreur type I de la tendance temporelle R, nous considérons :
- N : le nombre total de simulations,
- n : indice de la simulation,
- m_R : la vraie valeur de la tendance temporelle moyenne,
- $\hat{m}_R^n$, $n = 1, ..., N$: l'estimateur du paramètre m_R fourni par le modèle,
- $\hat{\sigma}_R^n$, $i = 1, ..., N$: l'estimateur de l'écart-type σ_R de l'estimation de m_R fourni par le modèle.

Considérons le test d'adéquation suivant :

$$\mathcal{H}_0 : m_R = m_R^{vrai} \textbf{ contre } \mathcal{H}_1 : m_R \neq m_R^{vrai}$$

L'erreur de type I associée au test d'adéquation est la probabilité de rejeter à tort l'hypothèse nulle :

$$ET_I = \mathbb{P}(rejeter \quad \mathcal{H}_0 | \mathcal{H}_0 \quad vraie).$$

La statistique de test est donnée par $T = \dfrac{\hat{m}_R - m_R}{\hat{\sigma}_R}$ et suit une loi de Student sous $\mathcal{H}_0$.

Pour l'estimation de l'erreur type I, nous utilisons l'approche par intervalle de confiance. En effet, l'erreur de type I correspond à la probabilité que l'intervalle de confiance fourni par le modèle ne contienne pas la vraie valeur du paramètre. Elle est estimée par le pourcentage d'intervalles de confiance fournis par le modèle qui ne contiennent pas la vraie valeur du paramètre. Alors, l'erreur de type I de la tendance temporelle est donnée par :

$$\hat{ET}_I = \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{\left(m_R \notin]\hat{m}_{R,inf}^n, \hat{m}_{R,sup}^n[\right)} \tag{3.15}$$

où $]\hat{m}_{R,inf}^n, \hat{m}_{R,sup}^n[$ est l'intervalle de confiance pour le paramètre m_R fourni par le modèle à la n_ième simulation. Les intervalles de confiance se calculent dans le cas fréquentiste sous l'hypothèse de normalité asymptotique des estimateurs. On fixe le seuil de significativité $\alpha = 5\%$. Les bornes de l'intervalle de confiance asymptotique du paramètre m_R au niveau de confiance $1 - \alpha$ sont donnée par

$$\begin{cases} \hat{m}_{R,inf} = \hat{m}_R - q\hat{\sigma}_R \\ \hat{m}_{R,sup} = \hat{m}_R + q\hat{\sigma}_R \end{cases}$$

où q est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi de Student approximée par une loi normale centrée réduite lorsque N est grand.

Chapitre 4

Résultats

Ce chapitre est consacré à la présentation des résultats issus des différents scénarios de simulations décrits dans le chapitre 3 ainsi que leurs interprétations. Mais dans un premier temps, on présente les estimateurs obtenus pour une simulation.

4.1 Approche simulation-ré-estimation

Dans cette partie, nous allons vérifier qualitativement sur un jeu de donnée simulé si le modèle proposé estime bien les paramètres de notre modèle. Le nombre de passages sur les placettes permanentes est égale à 12. La proportion de placettes temporaires est $p = 0.6$. Ainsi, nous disposons de 2000 placettes au total pour le suivi. La grille latente utilisée est de taille 15×15. Les résultats obtenus sont consignés dans le tableau (4.1).

	True value	Estimate	Std. Error	CI
$m_{\log(K^0)}$	1.000000000	1.24534128	0.19950753	[0.8543065, 1.636376]
m_R	0.040000000	0.04837195	0.02811199	[-0.00672755, 0.1034715]
ϕ	3.000000000	2.83496378	0.14087631	[2.558846, 3.111081]
$\log(\tau_u)$	-4.038101	-4.06669905	0.11949999	[-4.300919, -3.832479]
$\log(\tau_{\log(K^0)})$	-5.647539	-5.63581845	0.12196345	[-5.874867, -5.39677]
$\log(\tau_R)$	-3.750419	-3.87452600	0.13275364	[-4.134723, -3.614329]
$\log(\kappa)$	3.688879	3.71412439	0.07928754	[3.558721, 3.869528]
$logit(\rho)$	-1.386294	-1.35087392	0.07541696	[-1.498691, -1.203057]
$\log(\sigma_{\mu^0})$	-1.203973	-1.22582322	0.03290188	[-1.290311, -1.161336]
$\log(\sigma_{\mu^1})$	-1.609438	-1.61898282	0.02109915	[-1.660337, -1.577628]
$\log(\theta_0)$	.	-10.00000000	11.46871808	.
$\log(\theta_1)$	.	-10.00000000	2.20559979	.

TABLE 4.1 – Estimateurs et intervalles de confiance à 95%

Le tableau 4.1 montre que les estimateurs obtenus sont proches des vraies valeurs. Nous constatons que les vraies valeurs des transformations des paramètres se trouvent dans les intervalles de confiance à 95% calculés. On peut en déduire que le modèle estime bien les paramètres même

si on peut accepter que le vrai paramètre ne soit pas dans l'intervalle de confiance dans 5% des cas.

4.2 Résultats des scénarios de simulation

4.2.1 Résultats du scénario 1

Présentons d'abord les résultats du scénario 1 basés sur 1000 simulations. Dans ce présent rapport sont illustrés les boites à moustache des estimation de la tendance temporelle $\hat{m}_R$, l'incertitude de l'estimation $SE(\hat{m}_R)$ en fonction des plans d'échantillonnage considérés mais aussi de la quantité Δ^q_{SE}.

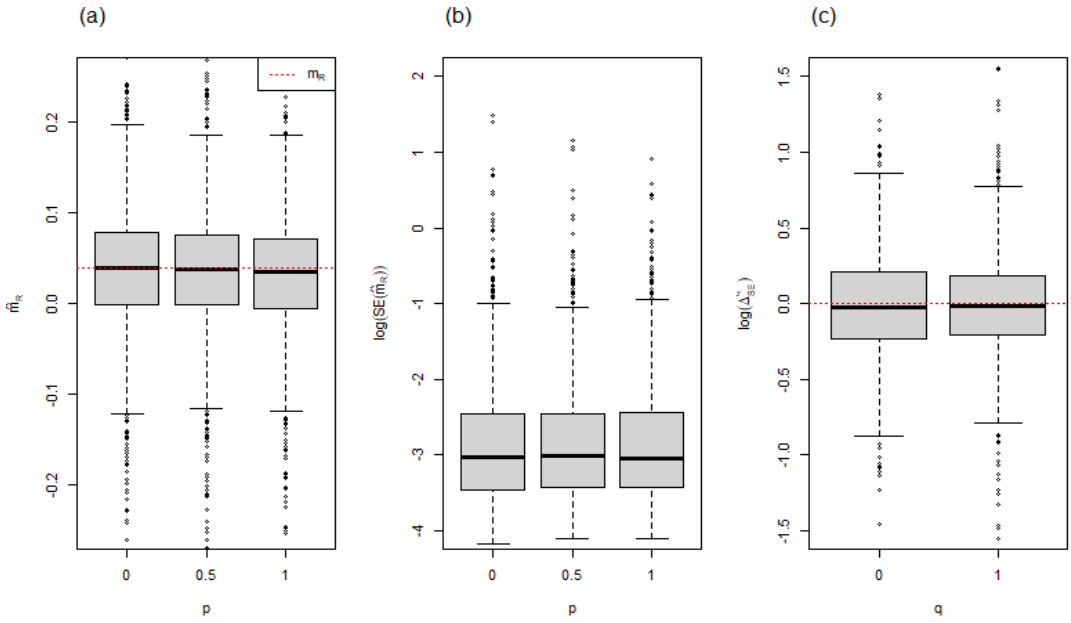

FIGURE 4.1 – Boxplot de l'estimation de la tendance temporelle, de l'incertitude de l'estimation et de Δ^q_{SE} en absence de nuggets

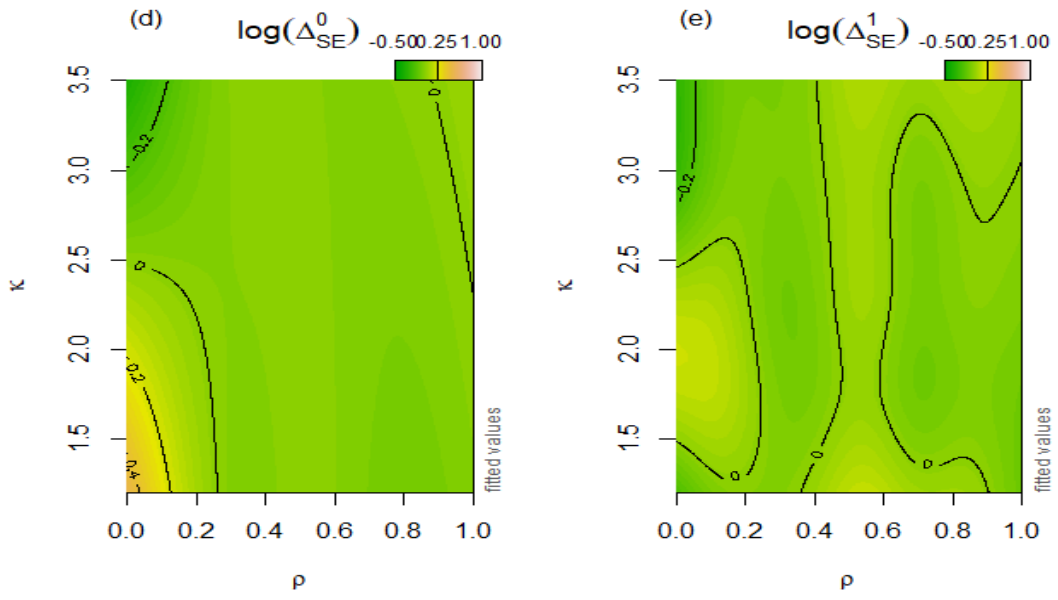

FIGURE 4.2 – Variation de $\log(\Delta^q_{SE})$ avec $q = 0, 1$, en fonction de ρ et κ en absence de nuggets.

Le premier point que l'on observe, à la vue de la figure (4.1), est que les 3 plans d'échantillonnages sont presque identiques en termes d'estimation de la tendance temporelle mais aussi de l'incertitude d'estimation. On observe que les médianes obtenues pour les plans d'échantillonnage sont proches de la vraie valeur de la tendance temporelle moyenne. On obtient presque les mêmes incertitudes d'estimation pour les 3 plans d'échantillonnage.

Nous nous intéressons à la variation de $\log(\Delta^q_{SE})$, $q = 0, 1$, en fonction de ρ et κ, représentée dans la figure (4.2). Au niveau du graphe (d), on remarque que :

— lorsque ρ est faible et κ élevé, le rapport $\Delta^0_{SE} = \dfrac{SE(\hat{m}_R)_0}{SE(\hat{m}_R)_{0.5}}$ est inférieur à 1. Ce qui implique dans ces conditions que l'incertitude de l'estimation de la tendance temporelle est plus faible pour un échantillonnage avec uniquement des placettes permanentes que pour un échantillonnage avec un mélange équilibré de placettes permanentes et temporaires. Donc la meilleure stratégie d'échantillonnage est un échantillonnage avec uniquement des placettes permanentes.

— De même, lorsque ρ et κ sont faibles, le rapport Δ^0_{SE} est supérieur à 1. Donc pour des valeurs du couple (ρ, κ) faibles, on obtient une incertitude plus faible avec un mélange équilibré de placettes permanentes et temporaires. Un échantillonage avec un mélange de placettes permanentes et temporaires est alors la meilleure stratégie d'échantillonnage pour l'estimation de la tendance temporelle.

— En dehors des valeurs du couple (ρ, κ) énumérées ci-dessus, les 2 plans d'échantillonnage sont identiques en terme de précision.

L'analyse du graphe (e) de la figure 4.2 montre que les plan d'échantillonnage avec uniquement de placettes temporaires et un mélange équilibré de placettes permanentes et temporaires sont identiques sauf pour ρ faible et κ fort. Dans ce cas, la meilleure stratégie est d'utiliser un échantillonnage avec un mélange équilibré de placettes permanentes et temporaires pour l'estimation de la tendance temporelle.

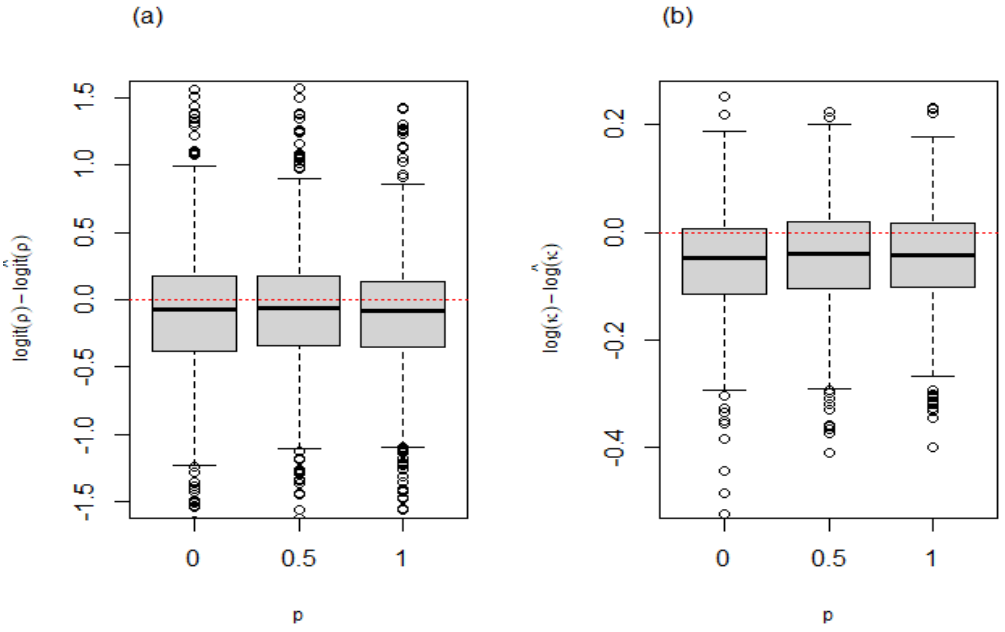

FIGURE 4.3 – Boxplot $logit(\rho) - \hat{logit}(\rho)$ et $log(\kappa) - \hat{\log}(\kappa)$ en absence de nuggets

On s'intéresse également à l'estimation des tranformées de ρ et κ. L'analyse du graphe (a) de la figure (4.3) montre que les 3 plans d'échantillonnage donnent des estimations presque identiques pour $logit(\rho)$ et $\log(\kappa)$. Le graphe (a) montre des estimations peu biaisées pour l'estimation de $logit(\rho)$ et le graphe (b) nous montre que les 3 plans d'échantillonnage fournissent des estimations fortement biaisées de $\log(\kappa)$.

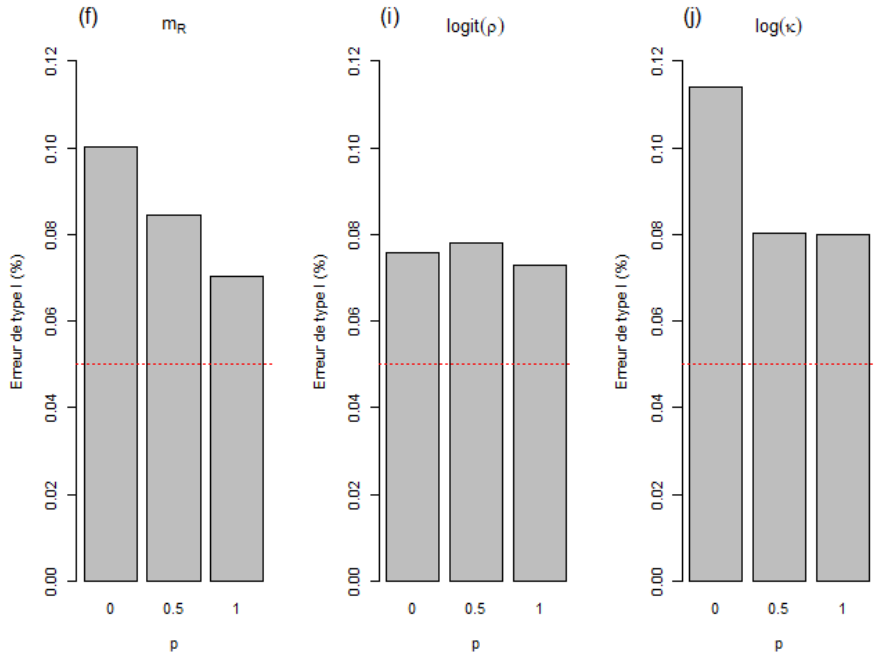

FIGURE 4.4 – Erreur de type I de m_R, $logit(\rho)$ et $\log(\kappa)$

L'analyse du graphe (i) de la figure (4.4) portant sur la qualité globale de l'inférence montre un problème d'adéquation de la méthode d'analyse suivant les plans d'échantillonnage et cela est visible sur la figure (4.4). Pour l'estimation de la tendance temporelle moyenne, l'échantillonnage avec uniquement de placettes permanentes donne la plus grande valeur en terme d'erreur de type I qui est de l'ordre de 10%. Alors que l'échantillonnage avec uniquement de placettes temporaires est associé une erreur de type I de l'ordre de 7%, la plus faible. En ce qui concerne l'estimation de la transformée de l'autocorrélation temporelle, les 3 plans d'échantillonnages donnent des erreurs de type I semblables. L'échantillonnage avec uniquement de placettes temporaires et un mélange de placettes permanentes et temporaires ont des erreurs de type I similaires pour la transformée du paramètre d'échelle alors que l'échantillonnage avec uniquement de placettes permanentes est associé à l'erreur de type I la plus grande.

4.2.2 Résultats du scénario 2

On fournit maintenant les résultats issus du scénario 2 qui correspond à l'introduction des effets nuggets dans le modèle. Les même types de graphiques sont réalisés.

Tout comme dans le scénario 1, on observe avec les graphes (a) et (b) de la la figure (4.5) que les estimations de la tendance temporelle moyenne et les incertitudes d'estimation sont presque identiques.

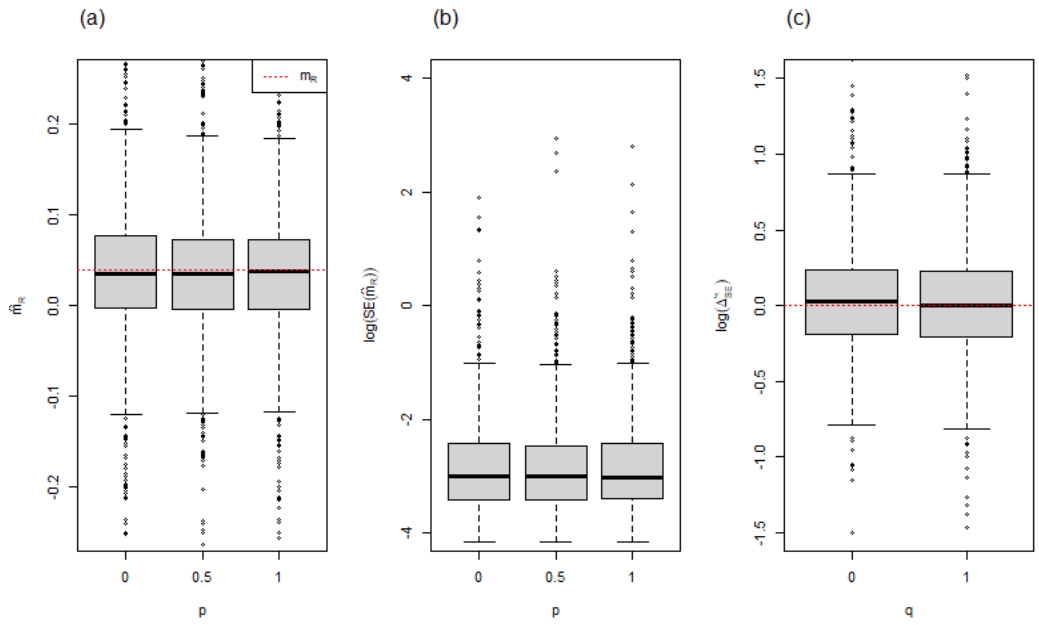

FIGURE 4.5 – Boxplot de l'estimation de la tendance temporelle, de l'incertitude de l'estimation et de Δ^q_{SE} en présence de nuggets

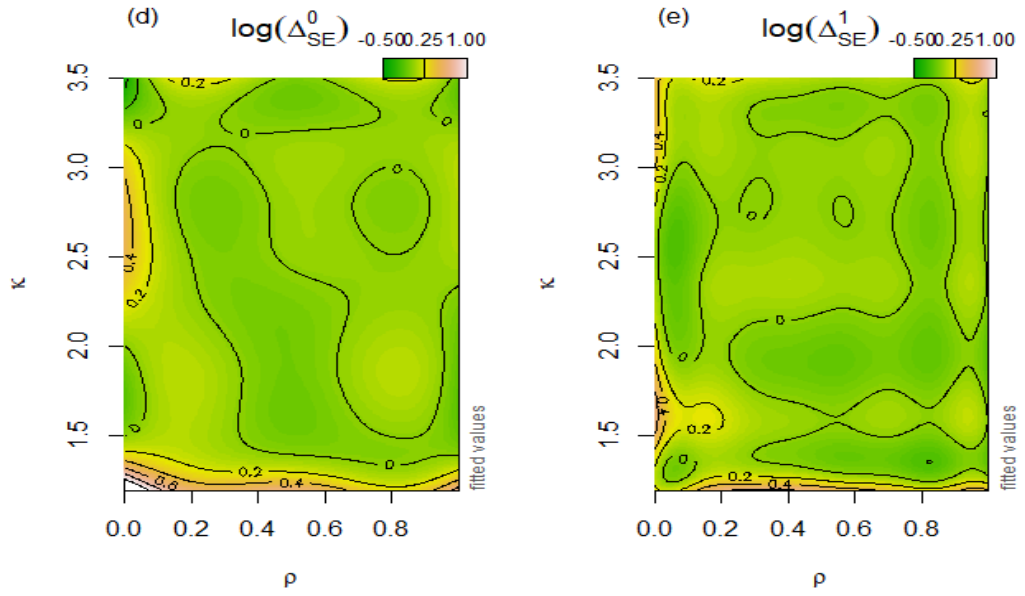

FIGURE 4.6 – Variation de $\log(\Delta^q_{SE}), q = 0, 1$, en fonction de ρ et κ en présence de nugget

Suivant les valeurs du couple (ρ, κ), on peut comparer les incertitudes de l'estimation de la tendance temporelle obtenues avec les plans d'échantillonnages avec $p = 0$ et $p = 1$ par rapport au plan d'échantillonnage avec $p = 0.5$. Au niveau de la figure (4.6), on remarque que :

— sur le graphe (d), pour des valeurs de κ faible et quelqe soit la valeur de ρ, le rapport Δ^0_{SE} est supérieur à 1. C'est aussi le cas pour ρ faible et κ moyen. L'échantillonnage avec un mélange équilibré de placettes permanentes et temporaires donne des incertitudes plus faibles que l'échantillonnage avec uniquement de placettes permanentes. Par conséquent, la meilleure stratégie d'échantillonnage pour l'estimation de la tendance temporelle est un échantillonnage avec un mélange équilibré de placettes permanentes et temporaires.

— sur le graphe (e), le rapport $\Delta^1_{SE} = \frac{SE(\hat{m}_R)_1}{SE(\hat{m}_R)_{0.5}}$ est supérieur à 1 dans les cas suivant : (1) κ faible et ρ moyen ou fort, (2) κ fort et ρ faible, (3) κ moyen et ρ faible. Donc

pour ces gammes de valeurs du couple (ρ, κ), l'échantillonnage avec un mélange équilibré de placettes permanentes et temporaire donne des estimations plus précis qu'un échantillonnage avec uniquement des placettes temporaires. Donc la meilleure stratégie d'échantillonnage pour l'estimation de la tendance temporelle est un échantillonnage avec un mélange de placettes permanentes et temporaires. En dehors de ces valeurs, les deux plans d'échantillonnage donnent les même précision sur l'estimation de la tendance temporelle moyenne.

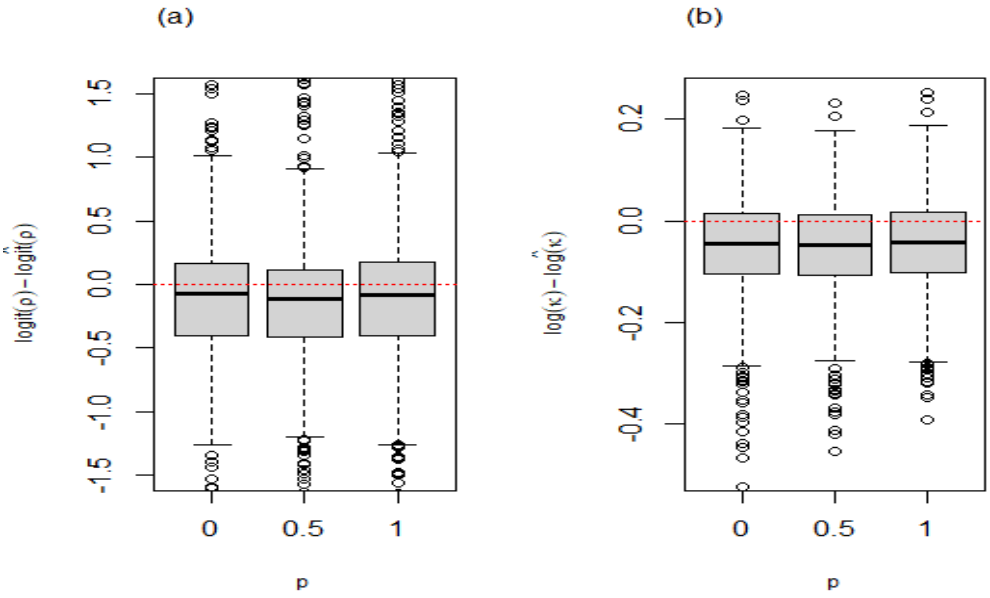

FIGURE 4.7 – Boxplot $logit(\rho) - \widehat{logit}(\rho)$ et $log(\kappa) - \widehat{\log}(\kappa)$ en présence de nuggets

L'analyse de la figure (4.7) montre que les estimations des transformées de ρ et κ obtenues avec les 3 plans d'échantillonnage sont presque identiques et biaisées.

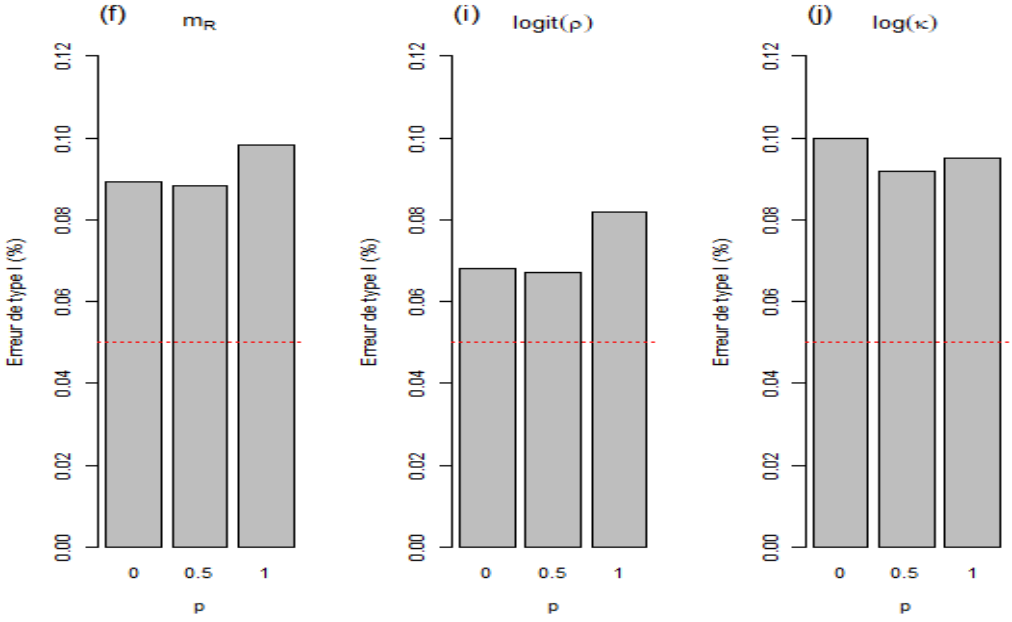

FIGURE 4.8 – Erreur de type I de m_R, $logit(\rho)$ et $\log(\kappa)$ en présence d'effet nuggets

Pour l'estimation de la tendance temporelle moyenne, au vue du graphe (f) de la figure (4.8), on observe peu de différence en terme d'erreur de type 1 entre l'échantillonnage avec uniquement de placettes permanentes et celui avec un mélange équilibré de placettes permanentes et temporaires. L'échantillonnage avec uniquement de placettes temporaires est associé l'erreur de type I la plus grande, de l'ordre de 9.8%.

4.2.3 Résultats du scénario 3

Présentons à présent les résultats liés à la comparaison des modèles statistiques.

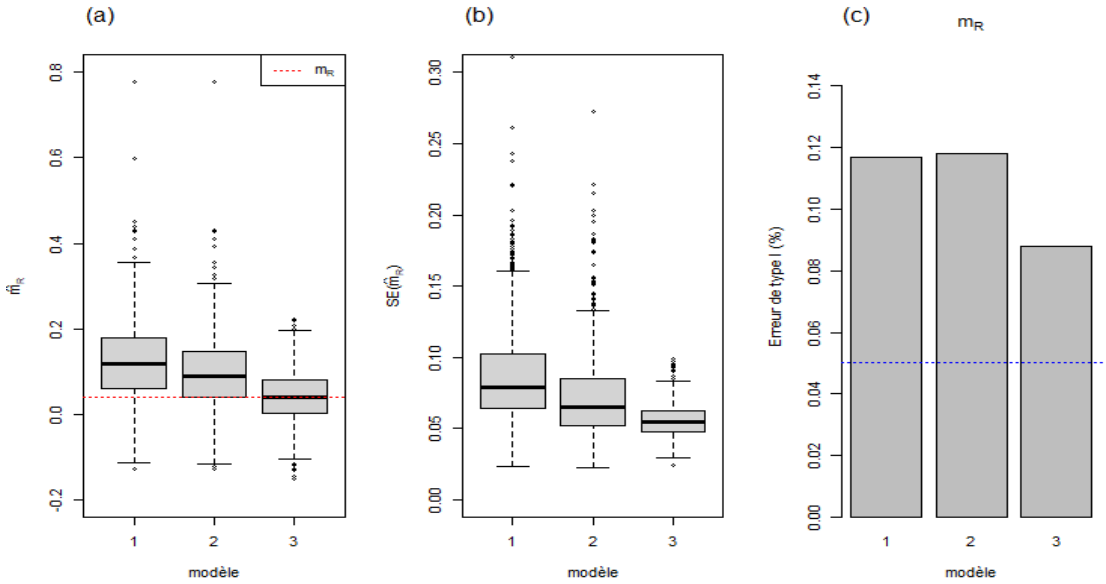

FIGURE 4.9 – Estimation, précision et erreurs de type I associées aux 3 modèles pour l'estimation de la tendance temporelle

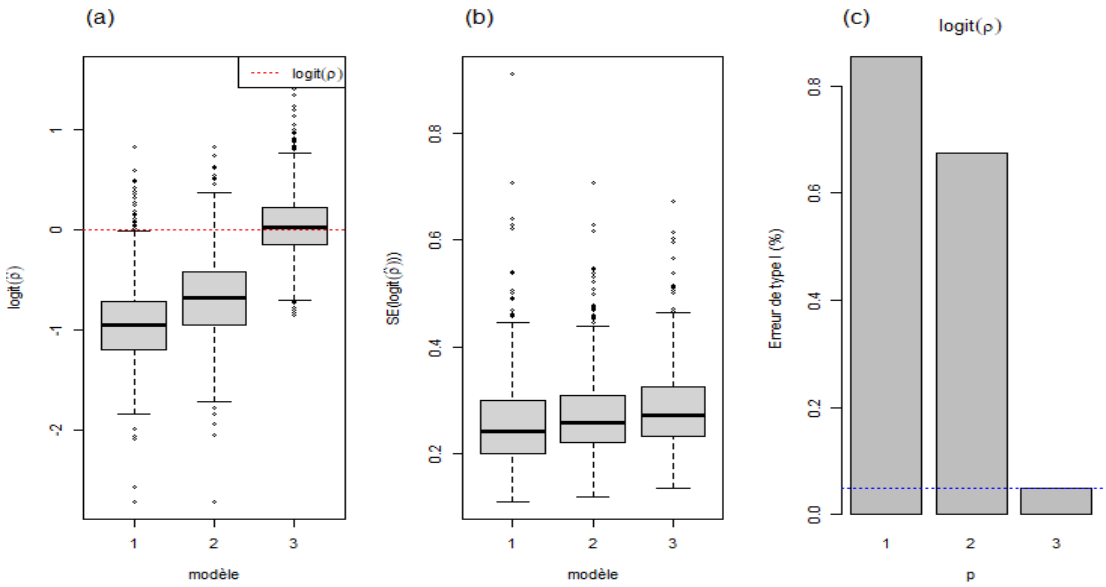

FIGURE 4.10 – Estimation, précision et erreurs de type I associées aux 3 modèles pour l'estimation de la transforme de ρ

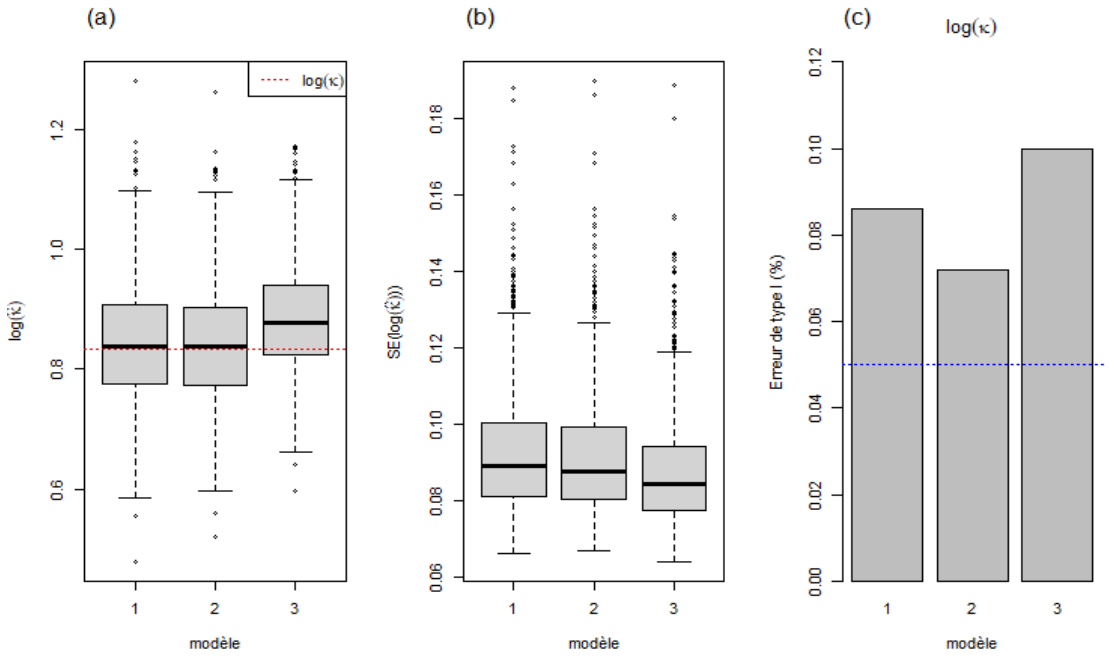

FIGURE 4.11 – Estimation, précision et erreurs de type I associées aux 3 modèles pour l'estimation de la transforme de κ

L'analyse des figures nous montre que :

— Le *modèle 1* est le plus mauvais modèle en terme d'estimation de la tendance temporelle moyenne et de précision (4.9). On remarque, à la vue du graphe (a) de la figure (4.9), que la médiane, qui se confond dans ce cas avec la moyenne car la distribution est approximativement normale, est largement au dessus de la vraie valeur de la tendance temporelle moyenne. Il est associé à une erreur de type I relativement grande, de l'ordre de 11.7%. Pour l'estimation de la transformée de ρ, le *modèle 1* fournit des estimations fournit des estimateurs fortement biaisés et peu précis et est associé à une erreur de type I trop grande, de l'ordre de 85% (figure 4.10). Cependant, il fournit des estimations non biaisées pour la transformée de κ.

— Le *modèle 2* est bien meilleur que le *modèle 1* en termes d'estimation de la tendance temporelle mais aussi de précision. Tout comme le *modèle 1*, le *modèle 2*, la médiane se trouve au dessus de la vraie valeur de la tendance temporelle moyenne et l'erreur de type I est de l'ordre 11.8% (4.9). Il fournit également des estimations très biaisé de $logit(\rho)$ associé à une erreur de type I de l'ordre 67.4%. Les estimations obtenues avec ce modèle pour $\log(\kappa)$ sont sans biais.

— Le *modèle 3* est le meilleur modèle implémenté pour l'estimation de la tendance temporelle. On observe que la médiane est environ égale à la vraie valeur de la tendance temporelle moyenne. Le *modèle 3* se différencie des *modèles 1* et *2* par une meilleure estimation de la tendance temporelle, une incertitude beaucoup plus petite mais aussi un erreur de type I plus faible. Le *modèle 3* est associée à une erreur de type I plus faible, de l'ordre 8.8%. Le *modèle 3* donne des estimations sans biais mais n'est pas parfait en termes d'erreur de type I. Concernant l'estimation de $logit(\rho)$, ce modèle fournit des estimations non biaisées mais moins précises et est parfait en terme d'erreur de type I. Néanmoins, les estimateurs pour $\log(\kappa)$ sont biaisés et l'erreur de type I obtenu est relativement grande de l'ordre 10%.

L'analyse des résultats montre que les *modèles 1* et *2* donnent des estimations fortement biaisées de la tendance temporelle moyenne et de l'autocorrélation spatiale alors que le *modèle 3* fournit des estimations approximativement non biaisées et une précision beaucoup plus faible des estimations. Le *modèle 3*, dont la tendance temporelle varie de manière spatiale et structurée, est le seul qui est a priori pertinent pour modéliser la tendance temporelle, car il est en mesure de modéliser des sources de variabilités additionnelles corrélées. On observe en effet, sur l'ensemble des simulations, que le *modèle 3* est globalement le plus performant, en termes d'inférence de la tendance temporelle et de précision réelle des estimations.

Chapitre 5

Discussion

5.1 Discussion sur les résultats

Les résultats des scénarios 1 et 2 nous ont montrés qu'il n'existe pas de différence significative globale entre les 3 plans d'échantillonnages considérés en terme d'estimation de la tendance temporelle et de précision. Cependant, on voit des différences en terme de précision des plans d'échantillonnage que dans des zones précises de l'espace (ρ, κ). Les conclusions à tirer diffèrent entre le scénario 1 et le scénario 2. Mais comme dans Rhodes & Jonzen, 2011 [16], il peut être préférable d'utiliser un type d'échantillonnage plutôt que d'autres suivant les valeurs de l'autocorrélation temporelle et du paramètre d'échelle lié à l'autocorrélation spatiale.

Les résultats nous montrent qu'en absence d'effet de nuggets (scénario 1), la meilleure stratégie d'échantillonnage est un échantillonnage avec uniquement des placettes permanentes ou uniquement des placettes temporaires plutôt qu'un échantillonnage avec mélange équilibré de placettes permanentes et temporaires. lorsque l'autocorrélation temporelle est faible et le paramètre d'échelle est élevé. En présence de nuggets dans les parties simulation et analyse statistique de données (scénario 2), on constate que la meilleure stratégie d'échantillonnage dans certaines régions de l'espace (ρ, κ) est le mélange équilibré de placettes permanentes et temporaires comparé à un échantillonnage avec uniquement de placettes permanentes ou uniquement de placettes temporaires. En effet, pour avoir un plan d'échantillonnage performant, il faut avoir un certain nombre de placettes revisitées plusieurs fois pour pouvoir bien estimer la part des nuggets.

A la différence de Rhodes et Jonzen 2011, nous avons utilisé un modèle statistique complet dans notre étude. Dans les travaux de Rhodes & Jonzen, des paramètres tels que le nombre de passages sur les placettes permanentes, le nombre total d'observations varient alors dans ce présent travail seule la proportion de placettes temporaires varie parmi les paramètres du plan d'échantillonnage. Cela peut expliquer pourquoi nous n'avons pas pu vérifier certains résultats de Rhodes & Jonzen dans un contexte statistique complet. Dans notre étude, sur une grande partie de l'espace (ρ, κ), il n'existe pas de différence nette entre les 3 plans d'échantillonnage. Houessou (2021 [15]) qui utilise les même méthodes que Rhodes & Jonzen, avait déjà trouvé que les différences entre les plans d'échantillonnage étaient faibles. De plus, on se base sur la quantité Δ_{SE}^{q} avec $q = 0, 1$ pour déterminer la meilleure stratégie d'échantillonnage. Ce qui n'est pas le cas dans les travaux de Rhodes & Jonzen. Il faut noter aussi que Rhodes & Jonzen utilise une gamme du nombre de repassage sur les placettes permanentes plus grande. Dans leur travail, ce paramètre prend les valeurs discrètes suivantes $\{2, 3, 4, 5, 10\}$ alors qu'il est fixé à 4 dans notre étude.

Les résultats de la comparaison des modèles pour l'estimation de la tendance temporelle (scénario 3) montrent qu'il est important de tenir compte de la corrélation entre les placettes dans la modélisation lorsque la tendance temporelle de l'espèce est bien structurée dans l'espace. Considérer des modèles où la tendance temporelle est constante ou varie aléatoirement dans l'espace et de manière non structurée conduisent à des estimations biaisées de la tendance temporelle et peu précises. De plus, ces modèles sont associés à des erreurs de type I relativement élevées. Les résultats montrent aussi que l'estimation de l'autocorrélation spatiale a une influence sur l'estimation de la tendance temporelle. Les deux premiers modèles statistiques considérés fournissent des estimations fortement biaisées de la tendance temporelle mais aussi de l'autocorrélation spatiale. La variance de la tendance temporelle est moyenne dans notre étude. Des simulations supplémentaires montrent que ces modèles donnent des estimation peu biaisées lorsque la variance de la tendance temporelle est faible. Ces biais augmentent lorsque la tendance temporelle a une variance plus grande (voir annexe 4.10 et 4.11). Ceci est en accord avec les résultats de Thorson et al. ,2015 [18] (qui a obtenu des résultats similaire pour l'estimation de la densité-dépendance), mais pour l'estimation de la tendance temporelle au lieu de l'estimation de la densité dépendance.

Les erreurs de type I calculées dans ce présent sont relativement grande. Cela peut s'expliquer par la précision de l'approximation SPDE utilisée ou l'approximation de Laplace utilisée par TMB dans l'estimation des paramètres. L'approximation SPDE a ses limites et il est important de mieux connaître ces limites, bien comprendre son impact sur l'erreur de type I des paramètres estimés.

Tous les résultats de ce présent travail sont basés sur une approche par simulation. Cette dernière est souvent utilisée pour explorer des questions écologiques qui ne peuvent être facilement traitées de manière empirique, et pour faire des prédictions sur l'avenir (Borcard et al. 2011). De même, les simulations sont utiles pour évaluer les méthodes d'estimation car les vraies valeurs des paramètres peuvent être prédéfinies dans les données de test. Les simulations peuvent également être utilisées pour comparer les options d'échantillonnage (Gruber et al. 2008).

5.2 Perspectives

En ce qui concerne le plan d'échantillonnage, les résultats de ce présent travail ne nous permettent pas de dégager un plan d'échantillonnage optimal pour tous les cas de figure mais plutôt la meilleure stratégie d'échantillonnage suivant certaines valeurs de l'autocorrélation temporelle et du paramètre d'échelle lié à l'autocorrélation spatiale. Il serait intéressent de se mettre dans le même cadre que Rhodes & Jonzen. Ainsi, on pourra faire varier le nombre de passages sur les placettes permanentes, le nombre total d'observations (budget) et d'autres paramètres du plan d'échantillonnage. Pour être beaucoup plus comparable, on pourrait aussi utiliser les méthodes de Rhodes & Jonzen et le modèle statistique porposé dans ce présent rapport sur les mêmes jeux de données avec les même méthodes d'analyse, notamment mettre l'accent sur l'étude de la quantité Δ^q_{SE} avec $q = 0, 1$.

Le modèle statistique utilisé semble très lié à la densité dépendance. Il serait intéressant d'utiliser des modèles avec d'autres formes de densité dépendance, voire des modèles sans densité dépendance et de refaire les analyses. Concernant l'inférence du modèle statistique complet pour l'estimation de la tendance temporelle, nous avons utilisé des estimations fréquentistes des paramètres du modèle. Il serait également intéressent de se mettre dans un cadre bayésien

pour l'estimation des paramètres. Coder le modèle statistique sous R-INLA peut être une bonne approche dans le cadre bayésien.

Chapitre 6

Conclusion

L'objectif de ce présent travail est de proposer un cadre statistique complet pour l'estimation de la tendance temporelle d'une population en se basant sur les travaux de Rhodes & Jonzen. Notre travail a principalement consisté à comparer les plans d'échantillonnage obtenus avec uniquement de placettes permanentes, uniquement de placettes temporaires et un mélange équilibré de placettes permanentes et temporaires pour l'estimation de la tendance temporelle d'une population. A la lumière des résultats obtenus pour la comparaison des plans d'échantillonnage, nous avons constaté qu'il n'existe pas une différence significative globale entre les 3 plans d'échantillonnage considérés. Cependant, avec ou sans effet de nuggets, un échantillonnage peut être considérée comme une meilleure stratégie d'échantillonage suivant les valeurs de l'autocorrélation temporelle et du paramètre d'échelle. Notre étude a montré qu'en présence de nuggets et pour certaines valeurs de l'autocorrélation temporelle et du paramètre d'échelle, un plan d'échantillonnage avec un mélange équilibré de placettes permanentes et temporaires est la meilleure stratégie d'échantillonnage pour k'estimation de la tendance temporelle. Nous ne sommes pas parvenus, dans cette étude, à généraliser toutes les conclusions de Rhodes & Jonzen (2011) dans un cadre statistique complet.

Nous avons ensuite comparé 3 modèles statistiques pour l'estimation de la tendance temporelle. Les performances des modèles ont été évaluées et comparées. D'après les résultats obtenus, lorsque la tendance temporelle de la population est bien structurée dans l'espace, le modèle dont la tendance temporelle varie dans l'espace de manière structurée donne les meilleurs résultats alors que les autres modèles se sont révélés être peu fiable pour l'estimation de la tendance temporelle moyenne dans notre étude. Dans la littérature d'écologie statistique, la tendance temporelle est modélisé de manière constante dans l'espace. Lorsque la tendance temporelle varie trop dans l'espace de manière structuré et que cette variation structuré dans l'espace n'est pas prise en compte dans la modélisation, les modèles fournissent des biais d'estimation considérables. Par conséquent, il est important de tenir compte de la corrélation entre les placettes dans la modélisation.

Ces résultats peuvent aider à concevoir de futurs plans d'échantillonnage dans le but de trouver le meilleur compromis entre une haute précision (maximiser la précision de l'estimation de la tendance temporelle) et le rapport coût-efficacité (définir un effort d'échantillonnage suffisant pour estimer la tendance temporelle avec précision tout en minimisant les coûts économiques).

Annexes

Annexe A

Introduction des strates

Nous supposons que l'espace des observations est subdivisé en J strates. De plus, nous considérons que la moyenne et la variance des champs aléatoires gaussiens $\log(K^0)$ et R diffèrent d'une strate à une autre. Différents modèles avec strates sont envisagés.

Modèle 1

Pour les différents noeuds de la grille latente, nous calculons la quantité $\log(N^0_{s,t})$, $\forall s$, donnée par

$$log(N^0_{s,t}) = \begin{cases} \phi - 0.5\sigma_u^2 + u_{s,t} & \text{si } t = 1 \\ (1-\rho)log(N^0_{s,t-1}) - 0.5\sigma_u^2 + u_{s,t} & \text{si } t > 1 \end{cases} \tag{A.1}$$

On modélise R et $\log(K^0)$ comme des champs aléatoires gaussiens de moyenne nulle mais avec une fonction de corrélation spatiale $M(.|\nu,\kappa)$ défini en (2.6)

$$R \sim MVN(0, M(.|\nu,\kappa))$$
$$\log(K^0) \sim MVN(0, M(.|\nu,\kappa))$$

Ces champs gaussiens sont générées sur la grille latente : donc de même taille que la grille latente.

Nous introduisons la variable $j(s) \in \{1, ..., s\}$ qui donne la strate dont appartient le noeud de la grille latente s. La log-abondance au niveau du noeud s est donnée par

$$log(N^{grid}_{s,t}) = \begin{cases} \log(N^0_{s,t}) + \left(m^{j(s)}_{\log(K^0)} + \sigma^{j(s)}_{\log(K^0)} \log(K^0_s)\right) & \text{si } t = 1 \\ \log(N^0_{s,t}) + \rho\left(\sum_{k=1}^{t-1}(1-\rho)^{t-1-k}\left(m^{j(s)}_{\log(K^0)} + \sigma^{j(s)}_{\log(K^0)} \log(K^0_s)\right)\right. & \\ \qquad \left. + \sum_{k=1}^{t-1} k(1-\rho)^{t-1-k}\left(m^{j(s)}_R + \sigma^{j(s)}_R R_s\right)\right) & \text{si } t > 1 \end{cases} \tag{A.2}$$

Comme dans le chapitre 3, nous déterminons les quatre noeuds de la grille latente entourant le site observé s et les coefficients des ces noeuds dans la pondération. La log-abondance au niveau du site observé r est donnée par A.3.

$$\log(N_{r,t}) = coef_{A_r} \log(N^{grid}_{A_r,t}) + coef_{B_r} \log(N^{grid}_{B_r,t}) + coef_{C_r} \log(N^{grid}_{C_r,t}) + coef_{D_r} \log(N^{grid}_{D_r,t}) \tag{A.3}$$

Il est important de noter que nous supposons que les noeuds A, B, C et D appartiennent à la même strate que le site observé r. Ce qui implique que $j(r) = j(A_r) = j(B_r) = j(C_r) = j(D_r)$

lors du calcul des quantités $\log(N_{s,t}^{grid})$, $s \in \{A_r, B_r, C_r, D_r\}$. Par cette modélisation, un noeud de la grille latente peut appartenir à plusieurs strates. Cela revient à calculer pour chaque sommet de la grille latente des $N_{s,t}^{grid}$ pour les différentes strates qui sont présentes autour du sommet.

Modèle 2

Nous pouvons aussi considérer que les strates sont définies uniquement pour les sites observés. Comme dans le modèle 1, les champs aléatoires gaussiens $\log(K^0)$ et R sont générés suivant la grille latente puis utilisés dans l'interpolation des observations. Contrairement au modèle 1, les hyperparamètres sont ceux de la strate où appartient le site observé. La formule d'interpolation est donnée par

$$log(N_{s,t}^{grid}) = \begin{cases} c_1 \log(N_{A,t}^0) + c_2 \log(N_{B,t}^0) + c_3 \log(N_{C,t}^0) + c_4 \log(N_{D,t}^0) \\ + \Big(m_{logK^0}^{j(r)} + \sigma_{logK^0}^{j(s)} \Big[c_1 \log(K_A^0) + c_2 \log(K_B^0) + c_3 \log(K_C^0) \\ + c_4 \log(K_D^0) \Big] \Big) & \text{si } t = 1 \\ \\ c_1 \log(N_{A,t}^0) + c_2 \log(N_{B,t}^0) + c_3 \log(N_{C,t}^0) + c_4 \log(N_{D,t}^0) \\ + \rho \Bigg(\sum_{k=1}^{t-1} (1-\rho)^{t-k-1} \Big(m_{logK^0}^{j(r)} + \sigma_{logK^0}^{j(s)} \Big[c_1 \log(K_A^0) \\ + c_2 \log(K_B^0) + c_3 \log(K_C^0) + c_4 \log(K_D^0) \Big] \Big) & \text{si } t > 1 \\ + \sum_{k=1}^{t-1} k(1-\rho)^{t-k-1} \Big(m_R^{j(r)} + \sigma_R^{j(r)} \Big[c_1 R_A + c_2 R_B + c_3 R_C + c_4 R_D \Big] \Big) \Bigg) \end{cases} \tag{A.4}$$

où $\log(N_{s,t}^0)$ est la quantité définie en (A.1)

Modèle 3

Un autre modèle peut être déduit des deux modèles précédents. Les noeuds de la grille latente appartiennent qu'à une seule strate. Donc pour les noeuds A_r, B_r, C_r et D_r, on calcule la quantité $\log(N_{s,t})$, $s \in \{A_r, B_r, C_r, D_r\}$,

Annexe B

Compléments sur les scénarios

B.0.1 Scénario 1

On présente les gams qui nous ont permis de représenter graphiquement la variation de Δ^q_{SE} avec $q = 0, 1$ en fonction de ρ et κ. Nous utilisons des modèles non linéaires en tenant compte de l'interaction de ρ et κ.

— La modèlisation de la variation de Δ^0_{SE} en fonction de ρ et κ par un gam donne le résultat suivant.

```
Family: gaussian
Link function: identity

Formula:
log(delta0) ~ s(rho) + s(kappa) + te(rho, kappa)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.01308    0.01344  -0.973    0.331

Approximate significance of smooth terms:
                edf Ref.df      F  p-value
s(rho)        1.000  1.000  7.597 0.005954 **
s(kappa)      1.000  1.000 11.127 0.000882 ***
te(rho,kappa) 5.604  6.472  3.923 0.001035 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.0234   Deviance explained = 3.09%
GCV = 0.18102  Scale est. = 0.17945   n = 994
```

Dans le modèle, on observe que ρ, κ ainsi que l'interaction de ρ et κ sont très significatives.

—
```
Family: gaussian
Link function: identity

Formula:
log(delta1) ~ s(rho) + s(kappa) + te(rho, kappa)

Parametric coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.006603   0.013081  -0.505    0.614
```

```
Approximate significance of smooth terms:
                edf Ref.df     F p-value
s(rho)        6.086  7.233 1.513   0.155
s(kappa)      1.000  1.000 0.130   0.719
te(rho,kappa) 10.048 13.047 1.470   0.129

R-sq.(adj) =  0.0213   Deviance explained = 3.83%
GCV = 0.17257  Scale est. = 0.16941   n = 990
```

Pour la modèlisation de Δ^1_{SE} en fonction de ρ et κ, On remarque que ρ, κ et l'interaction de deux ne sont pas significative

B.0.2 Scénario 2

Comme dans le scénario 1, on donne les résultats des modèles gam utilisés pour modéliser Δ^q_{SE} avec $q = 0, 1$ en fonction de ρ et κ en présence de nuggets.

—
```
Family: gaussian
Link function: identity

Formula:
log(delta0) ~ s(rho0, bs = "ps") + s(kappa0, bs = "ps") +
    te(rho0, kappa0, bs = "ps")

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.03334    0.01272   2.622  0.00889 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                  edf Ref.df     F p-value
s(rho0)         3.433  4.291 0.630 0.69277
s(kappa0)       8.348  8.802 2.542 0.00615 **
te(rho0,kappa0) 15.426 18.000 1.560 0.01296 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.0507   Deviance explained = 7.67%
GCV = 0.16513  Scale est. = 0.16043   n = 992
```

Pour modélisation de la variation de $\log(\Delta^0_{SE})$ en fonction de ρ et κ par un gam, on observe que κ et l'interaction de ρ et κ sont significatives alors que ρ n'est pas significative.

—
```
Family: gaussian
Link function: identity

Formula:
log(delta1) ~ s(rho1, bs = "ps") + s(kappa1, bs = "ps") +
    te(rho1, kappa1, bs = "ps")

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)  0.01623    0.01328   1.222    0.222

Approximate significance of smooth terms:
                    edf Ref.df     F p-value
s(rho1)           8.593  8.923 1.493  0.1043
s(kappa1)         8.059  8.513 3.044  0.0121 *
te(rho1,kappa1) 11.799 18.000 1.224  0.0218 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.0348   Deviance explained = 6.25%
GCV = 0.18023  Scale est. = 0.17488   n = 992
```

On constate les même remarque pour le modèle gam donnant Δ^1_{SE} en fonction de ρ et κ que pour Δ^0_{SE}.

B.0.3 Scénario 3

Dans le scénario 3, la variance de la tendance temporelle est de 0.035. On peut voir l'impact de cette variance sur la qualité de l'estimation de la tendance temporelle. Dans ce cas, nous considérons 1) une valeur faible de la variance de R, 2) et une valeur forte de la variance. Avec 1000 simulations pour chaque cas, on obtient les résultats suivant :

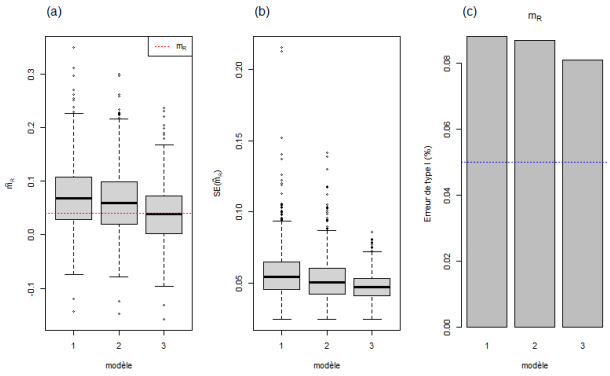

FIGURE B.1 – Résultat du scénario 3 avec $\sigma_{m_R} = 0.025$

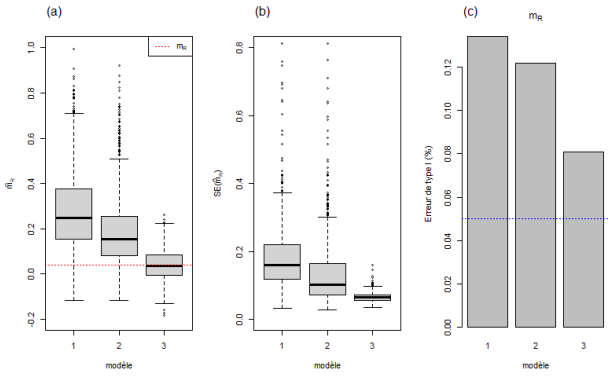

FIGURE B.2 – Résultat du scénario 3 avec $\sigma_{m_R} = 0.05$

D'après les résultats de la figure (B.1, les modèles 1 et 2 donnent des estimations peu biaisées de la tendance temporelle et sont associés à des erreurs de type I relativement grand lorsque la varaince de la tendance temporelle est faible. Mais lorsque cette variance est grande, ils fournissent des estimations fortement biaisées et peu précises (figure B.2).

On compare de plus les erreurs de type I des différents modèles avec un glm. On compare d'abord les erreurs de type I des modèles 1 et 2. On obtient le résultat suivant :

```
Call:
glm(formula = ech$Y ~ ech$type, family = binomial(link = "logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.4989  -0.4989  -0.4292  -0.4292   2.2047
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.02115    0.09838 -20.543   <2e-16 ***
ech$typeB   -0.31715    0.14879  -2.131    0.033 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1322.2  on 1999  degrees of freedom
Residual deviance: 1317.6  on 1998  degrees of freedom
AIC: 1321.6

Number of Fisher Scoring iterations: 5
```

Le coefficient $ech\$typeB$ donne la différence entre les erreurs de type I des deux modèles. On constate que les erreurs de type I sont significativement différent à 3.3%.

Ensuite, on compare les erreurs de type I des modèles 2 et 3.

```
Call:
glm(formula = ech$Y ~ ech$type, family = binomial(link = "logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.5011  -0.5011  -0.4292  -0.4292   2.2047

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.01151    0.09802  -20.52   <2e-16 ***
ech$typeB   -0.32680    0.14855   -2.20   0.0278 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1326.5  on 1999  degrees of freedom
Residual deviance: 1321.6  on 1998  degrees of freedom
AIC: 1325.6

Number of Fisher Scoring iterations: 5
```

On en déduit que les erreurs de type I des modèles 2 et 3 sont significativement différentes à 2.78%.

Enfin, on compare l'erreur de type I du modèle 3 (erreur de type I la plus faible) avec 0.05.

```
Call:
glm(formula = ech1$Y ~ 1, family = binomial(link = "logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.4989  -0.4989  -0.4989  -0.4989   2.0715
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.02115    0.09838  -20.54   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 721.81  on 999  degrees of freedom
Residual deviance: 721.81  on 999  degrees of freedom
AIC: 723.81

Number of Fisher Scoring iterations: 4
```

Pour cela, on calcule l'intervalle de confiance fourni par le modèle glm pour l'intercept puis on regarde si $logit(0.05)$ appartient à cet intervalle. L'intervalle donné par le modèle glm est $[-2.213975, -1.828325]$. Or $logit(0.05) = -2.944439$ n'appartient pas à cet intervalle. On en conclut que l'erreur de type I du modèle 3 est significativement différent de 0.05.

Annexe C

Code avec le logiciel R

```
##############################################################################
#####                          load library                              ######
##############################################################################
library(parallel)
library(INLA)
library(TMB)
require(optimx)

##############################################################################
#####              creation of the lists of scenarios                    #####
##############################################################################
simul_lisT<-function(B,tmax,T_p,T_np=1,p,pas,Sync=TRUE,McDonald=TRUE,x=1,y=2){
  # B    = total number of observations
  # tmax = maximum monitoring time
  # S    = Total number of sites
  # S_p  = Total number of permanents sites
  # S_np = Total number of non permanents sites
  # T    = the average number of visits per site
  # T_p  = the total number of visits per permanents sites
  # T_np = the total number of visits per non permanents sites
  # p    = the proportion of non permanents sites
  T <- round((1-p)*T_p + p)
  S_p <- round(B*(1-p)/T)
  S_np<-round(B*p/T)
  S <- S_p + S_np
  a=T_p%%pas  #to manage visits on permanents sites

  while( (pas==0)|(a!=0)){
    stop('T_p is not a multiple of pas or Pas is equal to 0. Please do check it.')
  }

  if (T_p>tmax) return(NA)
  ############################ Locations  ########################################
  nS <- S
  n <- floor(sqrt(nS)) + 2
  dx <- dy <- 1/(n-1)
  A <- NULL
```

```
for(j in 1:n){
  for(i in 1:n){
    Tmp = c("Latitude" = (j-1)*dx, "Longitude"= (i-1)*dy)
    A = rbind(A, Tmp)
  }
}
A = data.frame(A, row.names=NULL)
n.subgrid <- (n -1)^2
n.sample.station <- sample(1:n.subgrid,size=nS)
o <- order(n.sample.station)
n.sample.station <- n.sample.station[o]
cor <- matrix(1:nrow(A),nrow=n,ncol=n)

pas <- (A[2,]$Longitude - A[1,]$Longitude)
Loc <- NULL
for( i in n.sample.station){
  if(A[i,]$Latitude+pas <= 1 & A[i,]$Longitude+pas <= 1){
    max.Lat=A[i,]$Longitude+pas/2
    max.Long=A[i,]$Latitude+pas/2
    Tmp <- c("Latitude"=runif(1,min=A[i,]$Latitude,max=A[i,]$Latitude + pas),
             "Longitude"=runif(1,min=A[i,]$Longitude,max=A[i,]$Longitude + pas))
  }
  if(A[i,]$Latitude+pas > 1 & A[i,]$Longitude+pas <= 1 ){
    Tmp <- c("Latitude"=runif(1,min=A[i,]$Longitude ,max=A[i,]$Longitude + pas),
             "Longitude"=runif(1,min=A[i,]$Latitude-pas ,max=A[i,]$Latitude))
  }
  if(A[i,]$Latitude+pas <= 1 & A[i,]$Longitude+pas > 1 ){
    Tmp <- c("Latitude"=runif(1,min=A[i,]$Longitude-pas ,max=A[i,]$Longitude),
             "Longitude"=runif(1,min=A[i,]$Latitude ,max=A[i,]$Latitude+pas))
  }
  Loc = rbind(Loc, Tmp)
}
Loc = data.frame(Loc, row.names=NULL)
dx <- (A[2,]$Longitude - A[1,]$Longitude)
# plot
plot(A,pch=20,cex=0.4)
for(i in seq(-1,nrow(A),by=1)){
  abline(v=i*dx,h=i*dx)
}
points(Loc, pch=20, col=4, cex=1)
##############################################################################
# McDonald's notation
if(McDonald==TRUE){
  num_panel_p <- x+y
  n_ind_p <- round(S_p/num_panel_p)
  if (tmax>S_np){
    num_panel_np <- S_np
    n_ind_np <- round(S_np/num_panel_np)
  }else{
    num_panel_np <- tmax
```

```
  n_ind_np <- round(S_np/num_panel_np) + 1
}
panel <- as.list(rep(NA,num_panel_p+num_panel_np))

# Panel with plot permanent
# systematic sampling
for(i in 1:num_panel_p){
  a <- rep(NA,n_ind_p)
  for(j in 0:n_ind_p-1){
    a[j+1] <- j*num_panel_p + i
  }
  a <- subset(a,a<=S_p)
  panel[[i]] <- a
}
seq <- round(tmax/(x+y))
lisT <- list(rep(NA,tmax))
for(i in 1:num_panel_p){
  b <- rep(0,tmax)
  for (j in 1:seq) {
    for(k in 1:x){b[(j-1)*(x+y)+k+(i-1)] <- (j-1)*(x+y)+k+(i-1) }
  }
  if(length(b)>tmax) b <- b[-(tmax+1:length(b))]
  b <- subset(b,b!=0)
  b <- b[1:T_p]
  b <- b[!is.na(b)]
  lisT[[i]] <- b
}

# Panel with plot non permenant
# systematic sampling
for(i in (num_panel_p+1):(num_panel_np+num_panel_p)){
  a <- rep(NA,n_ind_np)
  for(j in 0:n_ind_np-1){
    a[j+1] <- j*num_panel_np + (i-num_panel_p) + S_p
  }
  a <- subset(a,a<=S)
  panel[[i]] <- a
}
ii <- 1
for(i in (num_panel_p+1):(num_panel_np+num_panel_p)){
  lisT[[i]] <- ii
  ii = ii+1
}
ii <- 0

DF = NULL
for(k in 1:(num_panel_p+num_panel_np)){
  for(s in panel[[k]]){
    for(t in lisT[[k]]){
      Tmp = c("Site"=s, "Year"=t)
```

```
        DF = rbind(DF, Tmp)
      }
    }
  }
  DF = cbind( DF, 'Longitude'=Loc[DF[,'Site'],1], 'Latitude'=Loc[DF[,'Site'],2] )
  DF = data.frame(DF, row.names=NULL)
  DF = DF[order(DF$Site),]

  # visualization
  Tab <- matrix(FALSE,nrow=num_panel_np+num_panel_p,ncol=tmax)
  for (i in 1:nrow(Tab)) {
    for(j in lisT[[i]]){
      Tab[i,j] <- TRUE
    }
  }
  image(Tab,xlab="panel",ylab="time",axes=FALSE)
  simul_list <- list("DF"=DF,"panel"=panel,"lisT"=lisT,"Lat"=Loc[,"x"],
                     "Long"=Loc[,"y"],"nS"=S,"nT"=tmax,"Nobs"=nrow(DF),
                     "Sp"=S_p,"Snp"=S_np,"Tp"=T_p,"num_PP"=num_panel_p,
                     "num_PNP"=num_panel_np)
  return(simul_list)
  ##############################################################################
}else{
  #vector of permanent sites
  vecTp<-1
  if (T_p>1) vecTp <- floor(seq(1,tmax,len=T_p))
  #vecTp <- floor(c(vecTp,seq(tmax/T_p,tmax-1,by=tmax/T_p)))
  lisT<-as.list(rep(NA,S))
  if(S_p!=0){
    for(i in 1:S_p){
      if(Sync==FALSE | Sync==F){
        a <- abs(floor(vecTp + sample(c(-tmax/T_p,0,tmax/T_p),length(vecTp),
                       replace=TRUE)))
        k <- which(a>tmax)
        a[k] <- sample(1:tmax,length(k),replace=FALSE)
        a[which(a==0)] <- sample(1:floor(tmax/T_p)+1,1)
        doublons <- which(duplicated(a))
        while(length(doublons)!=0){
          for (j in doublons) {
            a[j] <- sample(1:tmax,1)
            doublons <- which(duplicated(a))
          }
        }
        o <- order(a)
        a <- a[o]
      }else{
        a <- vecTp
      }
      lisT[[i]] <- a
    }
```

```
    }
    # non permenante plot
    if (S_np>0){
      for(i in S_p+1:S_np){
        listT[[i]]<-sample(1:tmax,1)
      }
    }
    DF = NULL
    for(s in 1:S){
      for(t in listT[[s]]){
        Tmp = c("Site"=s, "Year"=t)
        DF = rbind(DF, Tmp)
      }
    }
    DF = cbind( DF, 'Longitude'=Loc[DF[,'Site'],1], 'Latitude'=Loc[DF[,'Site'],2] )
    DF = data.frame(DF, row.names=NULL)

    # Visualization
    Tab <- matrix(FALSE,nrow=S,ncol=tmax)
    for (i in 1:S) {
      for(j in listT[[i]]){
        Tab[i,j] <- TRUE
      }
    }
    image(t(Tab),ylab="plot",xlab="time",axes=FALSE)
    simul_list <- list("DF"=DF,"listT"=listT,"Loc"=Loc,"nS"=S,
                       "nT"=tmax,"Nobs"=nrow(DF),"Sp"=S_p,"Snp"=S_np,"Tp"=T_p)
    return(simul_list)
  }
}
##############################################################################
#####                    latent grid in a square domain                  #####
##############################################################################
latent.grid <- function(domain,n){
  A1 <- domain[1,]; A2 <- domain[2,]
  A3 <- domain[3,]; A4 <- domain[4,]
  Lx <- abs(A2[1]-A1[1])
  Ly <- abs(A4[2]-A1[2])
  dx <- dy <- Lx/(n-1)
  DF <- NULL
  ii <- 1
  for(i in 1:n){
    for(j in 1:n){
      Tmp = c("Latitude" = A1[1] + (i-1)*dx, "Longitude"=A1[2] + (j-1)*dy)
      DF = rbind(DF, Tmp)
    }
  }
  DF = data.frame(DF, row.names=NULL)
  return(DF)
}
```

```r
##############################################################################
# indexes of nodes surrounding an observation and coefficient in the weighting #
##############################################################################
modeleObs2 <- function(noeuds,coord){
  # noeeuds est un Data.Frame qui stocke les noeuds de la grille et coord les
  #points IFN contenant 3 variables les dates des relevés
  # les longitudes et les latitudes, Z les elements simulé

  k=1
  while(k<nrow(noeuds)){

    if(noeuds[k,1]==noeuds[k+1,1]) k=k+1
    else  {pas <- abs(noeuds[k,1]-noeuds[k+1,1]) }; break
  }# connaissant la structure des noeuds, le pas de la grille s'obtient en
  # faisant la différence de 2 Longitudes
  # ou de 2 latitudes successivement différentes
  pas <- noeuds[2,"Longitude"] - noeuds[1,"Longitude"]
  coordsIFN <- as.matrix(coord)
  #xmin,ymin,xmax et ymax sont les extrémités des longitudes et latitudes de la carte

  xmin <- min(noeuds[,"Longitude"])  #xmin= minimum des longitudes,
  xmax <- max(noeuds[,"Longitude"])  #xmax=max des longitudes,
  ymin <- min(noeuds[,"Latitude"])  #ymin=min des llatitudes et
  ymax <- max(noeuds[,"Latitude"])  #ymax=max des latitudes.

  jmax = 1 + (xmax-xmin)/pas
  imax = 1 + (ymax-ymin)/pas
  corres1 <- matrix(NA,nrow = imax,ncol = jmax)
  # la matrice de correspondance des [i,j] à Zeta
  corres1 <- matrix(1:nrow(noeuds),ncol=sqrt(nrow(noeuds)),byrow=TRUE)

  M <- c()
  pt1 <- c()
  zeta <- NULL
  zeta.2 <- NULL
  coef <- NULL
  for(k in 1:nrow(coordsIFN)){
    # calcul des autres points entourants coordsIFN[i,]
    pt1 <- c(ymin,xmin) + floor((coordsIFN[k,]-c(ymin,xmin))/pas)*pas    #  indice du no

    #indices
    i1 <- 1 + floor((coordsIFN[k,"Latitude"]-ymin)/pas)
    j1 <- 1 + floor((coordsIFN[k,"Longitude"]-xmin)/pas)
    # en bas a gauche
    i2 <- 1 + floor((coordsIFN[k,"Latitude"]-ymin)/pas)
    j2 <- 1 + ceiling((coordsIFN[k,"Longitude"]-xmin)/pas)
    # en haut a gauche
    i3 <- 1 + ceiling((coordsIFN[k,"Latitude"]-ymin)/pas)
    j3 <- 1 + ceiling((coordsIFN[k,"Longitude"]-xmin)/pas)
    # en haut a droite
```

```
    i4 <- 1 + ceiling((coordsIFN[k,"Latitude"]-ymin)/pas)
    j4 <- 1 + floor((coordsIFN[k,"Longitude"]-xmin)/pas)
    # en bas a droite

    zeta1 <- corres1[i1,j1]
    zeta2 <- corres1[i2,j2]
    zeta3 <- corres1[i3,j3]
    zeta4 <- corres1[i4,j4]

    M <- (coordsIFN[k,]-pt1)/pas
    zeta <- rbind(zeta,c(zeta1,zeta2,zeta3,zeta4))# zeta2=zeta1+1; zeta3=zeta4+1
    # changement de repère comme origine pt1 et calcul des nouveaux coordonnées, dans
    coef <- rbind(coef,c((1- M[1])*(1- M[2]),M[2]*(1- M[1]), M[1]* M[2],M[1]*(1- M[2])))
  }
  return(list("ind"=zeta,"coef"=coef,"corre1"=corres1))
}

##############################################################################
####          simulation of data from the sampling plan                  ######
##############################################################################
Simulation <- function(Lt, kappa, SD_u=0.5, SD_logK0=0.5, SD_R=0.1, mean_logK0,
                       mean_R=0.002,sd_nugget0=0.001,sd_nugget1=0.001,rho=0.8, phi){

  DP <- Lt$DF
  Loc <- Lt$Loc
  n_stations <- Lt$nS
  n_years <- Lt$nT
  nobs <- Lt$Nobs
  S <- DP$Site
  T <- DP$Year
  range <- sqrt(8)/kappa
  domain <- matrix(cbind(c(0,1,1,0),c(0,0,1,1)),ncol=2)
  pas <- range/14
  n.grid <- round((dist(domain)[1]/pas)) + 1
  noeuds <- latent.grid(domain,n.grid)
  ind_coef <- modeleObs2(noeuds,Loc)
  ind <- ind_coef$ind
  coef <- ind_coef$coef

  # variance marginale
  tau_u <- 1/(2*sqrt(pi)*SD_u*kappa)
  tau_logK0 <- 1/(2*sqrt(pi)*SD_logK0*kappa)
  tau_R <- 1/(2*sqrt(pi)*SD_R*kappa)

  mesh = inla.mesh.create( noeuds, refine=FALSE, extend=-0.5 )

  # Spatial model
  B.kappa <- matrix(log(kappa),1,1)
  B.tau_u <- matrix(log(tau_u),1,1)
  model_u <- inla.spde2.matern(mesh=mesh,alpha=2,B.tau=B.tau_u,B.kappa=B.kappa)
```

```
B.tau_logK0 <- matrix(log(tau_logK0),1,1)
model_logK0 <- inla.spde2.matern(mesh=mesh,alpha=2,B.tau=B.tau_logK0,B.kappa=B.kappa)

B.tau_R <- matrix(log(tau_R),1,1)
model_R <- inla.spde2.matern(mesh=mesh,alpha=2,B.tau=B.tau_R,B.kappa=B.kappa)

Q_u <- inla.spde2.precision(model_u)
Q_logK0 <- inla.spde2.precision(model_logK0)
Q_R <- inla.spde2.precision(model_R)

# Simulate logK0 and R
logK0.grid = inla.qsample(n=1,Q_logK0) + mean_logK0
R.grid = inla.qsample(n=1,Q_R) + mean_R

# Simulate u
u.grid = array(NA, dim=c(mesh$n,n_years))
for(t in 1:n_years){
  u.grid[,t] = inla.qsample(n=1,Q_u)
}

# Calculate logN.grid
logN.grid = array(NA, dim=c(nrow(noeuds),n_years))
for(s in 1:nrow(noeuds)){
  logN.grid[s,1] = logK0.grid[mesh$idx$loc[s]] + phi
  for(t in 2:n_years){
    logN.grid[s,t] = (1-rho)*logN.grid[s,t-1] + rho*(logK0.grid[mesh$idx$loc[s]] +
          (t-1)*R.grid[mesh$idx$loc[s]]) - 0.5*SD_u^2 + u.grid[mesh$idx$loc[s],t]
  }
}

# Calculate logN
logN = array(NA, dim=c(n_stations,n_years))
for(s in 1:n_stations){
  N = ind[s,]
  s1 = N[1];s2 = N[2];s3= N[3];s4= N[4];
  for(t in 1:n_years){
    logN[s,t] = coef[s,1]*logN.grid[s1,t] + coef[s,2]*logN.grid[s2,t] +
                coef[s,3]*logN.grid[s3,t] + coef[s,4]*logN.grid[s4,t]
  }
}

# nugget
nugget0 <- rnorm(n=n_stations, mean=0, sd=sd_nugget0)
nugget1 <- rnorm(n=n_stations, mean=0, sd=sd_nugget1)

# Simulate data
DF = NULL; N = NULL; c = NULL
for(o in 1:nobs){
  Tmp <- c( "Site"=S[o], "Year"=T[o], "Y.nugget"=rpois(1,lambda=exp(logN[S[o],T[o]] +
```

```
      nugget0[S[o]] + (T[o]-1)*nugget1[S[o]])), "Y"=rpois(1,lambda=exp(logN[S[o],T[o]] )
    DF = rbind(DF, Tmp)
  }
  for(o in 1:nobs){
    N = rbind(N,c("neighbour1"=ind[S[o],1],"neighbour2"=ind[S[o],2],
                  "neighbour3"= ind[S[o],3], "neighbour4"=ind[S[o],4]))
    c = rbind(c,c("coef1"=as.double(coef[S[o],1]),"coef2"=as.double(coef[S[o],2]),
                  "coef3"=as.double(coef[S[o],3]),"coef4"=as.double(coef[S[o],4])))
  }
  DF = data.frame(DF, row.names=NULL)
  N = data.frame(N, row.names=NULL)
  c = data.frame(c, row.names=NULL)

  # Return stuff
  Sim_List = list("DF"=DF, "N"=N,"Loc"=Loc,"logN"=logN,"noeuds"=noeuds,"n_years"=n_years
                  "n_stations"=n_stations,"n.grid"=n.grid^2,"coef"=c,"mesh"=mesh)
  Sim_List[["Parameters"]] = c("range"=range,'kappa'=kappa, "tau_u"=tau_u,
                               "tau_logK0"=tau_logK0,"tau_R"=tau_R,'Sigma_u'=SD_u,
                               'Sigma_logK0'=SD_logK0,'Sigma_R'=SD_R, 'rho'=rho,
                               "m_logK0"=mean_logK0,"m_R"=mean_R, "phi"=phi,
                               "sd_nugget0"=sd_nugget0,"sd_nugget1"=sd_nugget1)
  return(Sim_List)
}

###################
#  TMB functions
###################
RJ_GMRF_grid_nugget <- "
  // Space time
  #include <TMB.hpp>
  // Function for detecting NAs
  template<class Type>
  bool isNA(Type x){
    return R_IsNA(asDouble(x));
  }

  template<class Type>
  Type objective_function<Type>::operator() ()
  {
  // Indices
  DATA_INTEGER( n_obs );         // Total number of observations
  DATA_INTEGER( n_noeuds );         // Number of noeuds
  DATA_INTEGER( n_years );          // Number of years

  // Data
  DATA_IVECTOR( x_s ); // Association of each station with a given vertex in SPDE mesh
  DATA_VECTOR( Y );     // Count data
  DATA_IVECTOR( A );
  DATA_IVECTOR( B );
  DATA_IVECTOR( C );
```

```
DATA_IVECTOR( D );
DATA_VECTOR( coef1 );
DATA_VECTOR( coef2 );
DATA_VECTOR( coef3 );
DATA_VECTOR( coef4 );
DATA_IVECTOR( S );        // Station for each sample
DATA_IVECTOR( T );        // Time for each sample

// SPDE objects
DATA_SPARSE_MATRIX(G0);
DATA_SPARSE_MATRIX(G1);
DATA_SPARSE_MATRIX(G2);

// Fixed effects
PARAMETER(m_logK0);               // mean of Omega0
PARAMETER(m_R);                 // mean of Omega1
PARAMETER(phi);
PARAMETER(log_tau_u);             // log-inverse SD of u
PARAMETER(log_tau_logK0);           // log-inverse SD of logK0
PARAMETER(log_tau_R);            // log-inverse SD of R
PARAMETER(log_kappa);            // Controls range of spatial variation
PARAMETER(logitrho);             // Autocorrelation (i.e. density dependence)
PARAMETER(logSD_nugget0);      // log SD of nugget0
PARAMETER(logSD_nugget1);      // log SD of nugget1
PARAMETER_VECTOR(ln_VarInfl);   // Overdispersion parameters

// Random effects
PARAMETER_ARRAY(u_input);  // Spatial process variation
PARAMETER_VECTOR(logK0_input); // variation in carrying capacity
PARAMETER_VECTOR(R_input);
PARAMETER_VECTOR(nugget0);
PARAMETER_VECTOR(nugget1);

// objective function -- joint negative log-likelihood
using namespace density;
Type res = 0;
vector<Type> res_comp(3);
res_comp.setZero();

// Spatial parameters
Type kappa = exp(log_kappa);
Type tau_u = exp(log_tau_u);
Type tau_logK0 = exp(log_tau_logK0);
Type tau_R = exp(log_tau_R);
Type SD_nugget0 = exp(logSD_nugget0);
Type SD_nugget1 = exp(logSD_nugget1);
Type rho = 1/(1+exp(-logitrho));
Type kappa2 = exp(2.0*log_kappa);
Type kappa4 = kappa2*kappa2;
Type pi = 3.141592;
```

```
Type Range = sqrt(8) / exp( log_kappa );
Type Sigma_u = 1 / sqrt(4*pi*exp(2*log_tau_u)*exp(2*log_kappa));
Type Sigma_logK0 = 1 / sqrt(4*pi*exp(2*log_tau_logK0)*exp(2*log_kappa));
Type Sigma_R = 1 / sqrt(4*pi*exp(2*log_tau_R)*exp(2*log_kappa));
Eigen::SparseMatrix<Type> Q = kappa4*G0 + Type(2.0)*kappa2*G1 + G2;

// Objects for derived values
vector<Type> logK0(n_noeuds);
vector<Type> R(n_noeuds);
array<Type>  u(n_noeuds, n_years);

// Probability of Gaussian-Markov random fields (GMRFs)
res_comp(0) += GMRF(Q)(logK0_input);
res_comp(1) += GMRF(Q)(R_input);
res_comp(2) -= sum(dnorm(nugget0,0,SD_nugget0,true));
res_comp(2) -= sum(dnorm(nugget1,0,SD_nugget1,true));

for(int t=1; t<n_years+1; t++){
  res_comp(2) += GMRF(Q)(u_input.col(t-1));
}
// Transform GMRFs
array<Type> logN(n_noeuds,n_years);
for(int s=0; s<n_noeuds; s++){
  logK0(s) = m_logK0 + logK0_input(x_s(s))/exp(log_tau_logK0);
  R(s) = m_R + R_input(x_s(s))/exp(log_tau_R);
  for( int t=0; t<n_years; t++){
    u(s,t) = u_input(x_s(s),t)/exp(log_tau_u);
    // Calculate of logN in the grid
    if(t==0){logN(s,t) = logK0(s) + phi ;}
    if(t>=1){logN(s,t) = (1-rho)*logN(s,t-1) + rho*( logK0(s) + t*R(s) ) -
                         0.5*Sigma_u*Sigma_u + u(s,t);}
  }
}

// Likelihood contribution from observations
vector<Type> log_Yhat(n_obs);
for (int i=0; i<n_obs; i++){
  log_Yhat(i) = coef1(i)*logN(A(i),T(i)) + coef2(i)*logN(B(i),T(i) ) +
                coef3(i)*logN(C(i),T(i)) + coef4(i)*logN(D(i),T(i));
  log_Yhat(i) += nugget0(S(i)) + T(i)*nugget1(S(i)) ;
  Type mean_y = exp( log_Yhat(i) );
  Type var_y = mean_y*(1.0+exp(ln_VarInfl(0)))+pow(mean_y,2.0)*exp(ln_VarInfl(1));
  if( !isNA(Y(i)) ){
    res_comp(2) -= dnbinom2( Y(i), mean_y, var_y, 1 );
  }
}

res = res_comp.sum();

// Diagnostics
```

```
  REPORT( res_comp );     REPORT( res );
  // Spatial field summaries
  REPORT( Range );          REPORT( Sigma_u );
  REPORT( Sigma_logK0 );    REPORT( Sigma_R );    REPORT( kappa );
  REPORT( rho );            REPORT ( m_logK0 );   REPORT( tau_u );
  REPORT ( phi );           REPORT ( m_R );       REPORT( tau_logK0 );
  ADREPORT( Range );        ADREPORT( Sigma_u );  REPORT( tau_R );
  ADREPORT( Sigma_logK0 ); ADREPORT( Sigma_R );
  // Fields
  REPORT( u );      REPORT( logK0 );
  REPORT( R );      REPORT( logN );
  REPORT(SD_nugget0);
  REPORT(SD_nugget1);

  return res;
}"
# Compile
cat(RJ_GMRF_grid_nugget,file="RJ_GMRF_grid_nugget.cpp")
compile( "RJ_GMRF_grid_nugget.cpp",flags="-Wl,-V &> results_V0.txt")
dyn.load(dynlib("RJ_GMRF_grid_nugget"))

RJ_GMRF_grid <- "
  // Space time
  #include <TMB.hpp>
  // Function for detecting NAs
  template<class Type>
  bool isNA(Type x){
    return R_IsNA(asDouble(x));
  }

  template<class Type>
  Type objective_function<Type>::operator() ()
  {
  // Indices
  DATA_INTEGER( n_obs );         // Total number of observations
  DATA_INTEGER( n_noeuds );         // Number of noeuds
  DATA_INTEGER( n_years );          // Number of years

  // Data
  DATA_IVECTOR( x_s );// Association of each station with a given vertex in SPDE mesh
  DATA_VECTOR( Y );                // Count data
  DATA_IVECTOR( A );
  DATA_IVECTOR( B );
  DATA_IVECTOR( C );
  DATA_IVECTOR( D );
  DATA_VECTOR( coef1 );
  DATA_VECTOR( coef2 );
  DATA_VECTOR( coef3 );
  DATA_VECTOR( coef4 );
  DATA_IVECTOR( S );        // Station for each sample
```

```
DATA_IVECTOR( T );        // Time for each sample

// SPDE objects
DATA_SPARSE_MATRIX(G0);
DATA_SPARSE_MATRIX(G1);
DATA_SPARSE_MATRIX(G2);

// Fixed effects
PARAMETER(m_logK0);                // mean of Omega0
PARAMETER(m_R);                 // mean of Omega1
PARAMETER(phi);
PARAMETER(log_tau_u);             // log-inverse SD of u
PARAMETER(log_tau_logK0);           // log-inverse SD of logK0
PARAMETER(log_tau_R);            // log-inverse SD of R
PARAMETER(log_kappa);            // Controls range of spatial variation
PARAMETER(logitrho);             // Autocorrelation (i.e. density dependence)
PARAMETER_VECTOR(ln_VarInfl);    // Overdispersion parameters

// Random effects
PARAMETER_ARRAY(u_input);  // Spatial process variation
PARAMETER_VECTOR(logK0_input); // variation in carrying capacity
PARAMETER_VECTOR(R_input);

// objective function -- joint negative log-likelihood
using namespace density;
Type res = 0;
vector<Type> res_comp(3);
res_comp.setZero();

// Spatial parameters
Type kappa = exp(log_kappa);
Type tau_u = exp(log_tau_u);
Type tau_logK0 = exp(log_tau_logK0);
Type tau_R = exp(log_tau_R);
Type rho = 1/(1+exp(-logitrho));
Type kappa2 = exp(2.0*log_kappa);
Type kappa4 = kappa2*kappa2;
Type pi = 3.141592;
Type Range = sqrt(8) / exp( log_kappa );
Type Sigma_u = 1 / sqrt(4*pi*exp(2*log_tau_u)*exp(2*log_kappa));
Type Sigma_logK0 = 1 / sqrt(4*pi*exp(2*log_tau_logK0)*exp(2*log_kappa));
Type Sigma_R = 1 / sqrt(4*pi*exp(2*log_tau_R)*exp(2*log_kappa));
Eigen::SparseMatrix<Type> Q = kappa4*G0 + Type(2.0)*kappa2*G1 + G2;

// Objects for derived values
vector<Type> logK0(n_noeuds);
vector<Type> R(n_noeuds);
array<Type>  u(n_noeuds, n_years);

// Probability of Gaussian-Markov random fields (GMRFs)
```

```
  res_comp(0) += GMRF(Q)(logK0_input);
  res_comp(1) += GMRF(Q)(R_input);

  for(int t=1; t<n_years+1; t++){
    res_comp(2) += GMRF(Q)(u_input.col(t-1));
  }
  // Transform GMRFs
  array<Type> logN(n_noeuds,n_years);
  for(int s=0; s<n_noeuds; s++){
    logK0(s) = m_logK0 + logK0_input(x_s(s))/exp(log_tau_logK0);
    R(s) = m_R + R_input(x_s(s))/exp(log_tau_R);
    for( int t=0; t<n_years; t++){
      u(s,t) = u_input(x_s(s),t)/exp(log_tau_u);
      // Calculate of logN in the grid
      if(t==0){logN(s,t) = logK0(s) + phi ;}
      if(t>=1){logN(s,t) = (1-rho)*logN(s,t-1) + rho*( logK0(s) + t*R(s) ) -
                           0.5*Sigma_u*Sigma_u + u(s,t);}
    }
  }

  // Likelihood contribution from observations
  vector<Type> log_Yhat(n_obs);
  for (int i=0; i<n_obs; i++){
    log_Yhat(i) = coef1(i)*logN(A(i),T(i)) + coef2(i)*logN(B(i),T(i) ) +
                  coef3(i)*logN(C(i),T(i)) + coef4(i)*logN(D(i),T(i));
    Type mean_y = exp( log_Yhat(i) );
    Type var_y = mean_y*(1.0+exp(ln_VarInfl(0)))+pow(mean_y,2.0)*exp(ln_VarInfl(1));
    if( !isNA(Y(i)) ){
      res_comp(2) -= dnbinom2( Y(i), mean_y, var_y, 1 );
    }
  }

  res = res_comp.sum();

  // Diagnostics
  REPORT( res_comp );     REPORT( res );
  // Spatial field summaries
  REPORT( Range );         REPORT( Sigma_u );
  REPORT( Sigma_logK0 );   REPORT( Sigma_R );    REPORT( kappa );
  REPORT( rho );           REPORT ( m_logK0 );   REPORT( tau_u );
  REPORT ( phi );          REPORT ( m_R );       REPORT( tau_logK0 );
  ADREPORT( Range );       ADREPORT( Sigma_u );  REPORT( tau_R );
  ADREPORT( Sigma_logK0 ); ADREPORT( Sigma_R );
  // Fields
  REPORT( u );      REPORT( logK0 );
  REPORT( R );      REPORT( logN );

  return res;
}"
# Compile
```

```
cat(RJ_GMRF_grid,file="RJ_GMRF_grid.cpp")
compile( "RJ_GMRF_grid.cpp",flags="-Wl,-V &> results_V0.txt")
dyn.load(dynlib("RJ_GMRF_grid"))
```

Annexe D

Appendice

L'accès aux stations de calcul de l'université d'Orléans et de l'université de Sorbonne ont considérablement facilité la mise en oeuvre des simulations. Chaque scénario a été réalisé avec 1000 jeux de données. Les simulations finales ont néanmoins nécessité environ un mois de calcul. La parallélisation des codes via les packages doMC et Parallel du logiciel R a largement contribuée à la réduction du temps de calcul qui aurait pu prendre plus deux mois.

Durant ce stage, j'ai été en contact avec Houessou, qui doit débuter sa thèse en écologie statistique au Canada au mois de Janvier. Son sujet d'étude s'intitule : "Modélisation statistique des mouvements migratoires des oiseaux en Amérique du Nord".

Ce stage va vraisemblablement donner lieu à la soumission d'un article en anglais dans une revue scientifique d'écologie statistique.

Enfin, le présent rapport a été rédigé en LATEX, qui est un langage déstiné à l'écriture de compositions scientifiques.

Bibliographie

[1] Amel Meddad-Hamza., - *Stratégie d'Echantillonnage en Ecologie.* -, support de cours, 2017.

[2] Asri A., Benamirouche R. - *Using INLA/SPDE Approach for Estimating a Spatial Model for Lung Cancer Mortality in Algeria 2016.* - journal of economics and applied statistics, vol. 18, n° 1, 2021, pp. 261-277.

[3] Auger-Méthé M., K. Newman, D. Cole, F. Empacher, R. Gryba, A. A. King, V. Leos-Barajas, J. Mills Flemming, A. Nielsen, G. Petris, and L. Thomas. 2021. - *A guide to state–space modeling of ecological time series.*- Ecological Monographs 91(4) :e01470. 10.1002/ecm.1470

[4] Asri A., Benamirouche R. - *Using INLA/SPDE Approach for Estimating a Spatial Model for Lung Cancer Mortality in Algeria 2016.* - journal of economics and applied statistics, vol. 18, n° 1, 2021, pp. 261-277.

[5] Carvalho S.B., Gonçalves J., Guisan A., Honrado J.P. - *Systematic site selection for multispecies monitoring networks.* - Journal of Applied Ecology, vol. 53, n° 5, 2016, pp. 1305-1316.

[6] Corona P., Fattorini L., Franceshi S., Scrinzi G., Torresan C. - *Estimation of standing wood volume in forest compartments by exploiting airborne laser scanning information : Model-based, Design-based, And hybrid perspectives.* - Canadian Journal of Forest Research, vol. 44, n° 11, 2014, pp. 1303-1311.

[7] Desassis N., Carrizo V.R and Allard D. - *Geostatistic with stochastic partial differential equations (SPDE)* - 2022.

[8] Edwards Jr T.C., Cutler D.R., Zimmermann N.E., Geiser L., Algria J. - *Model-based stratifications for enhancing the detection of rare ecological events.* - Ecology, vol. 86, n° 5, 2005, pp. 1081-1090.

[9] Elias Krainski, Virgilio G. Rubio, Haakon Bakka, Amanda Lenzi, Daniela C. Camilo, Daniel Simpson, Finn Lindgren and Håvard Rue. - *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA.*

[10] Finn Lindgren, Håvard Rue, and Johan Lindström. - *An explicit link between Gaussian fields and Gaussian Markov random fields : the stochastic partial differential equation approach.* - 2018, Chapman & Hall/CRC

[11] Gosselin F., Saas Y., - *Comparison of regression methods for spatially-autocorrelated count data on regularly- and irregularly-spaced locations.* - Ecography 37 : 476–489, 2014, `https://doi.org/10.1111/j.1600-0587.2013.00279.x`

[12] Levy P.S., and Lemeshow S., - *Sampling of Populations : Methods and Applications* -, New York : John Wiley Sons, 1999.

[13] Lindén A. and Mäntyniemi S., - *Using the negative binomial distribution to model overdispersion in ecological count data.* -Ecology, Volume 92, 2011, p. 1414-1421 `https://doi.org/10.1890/10-1831.1`

[14] McDonald, T.L., 2003. *Review of environmental monitoring methods : Survey designs. Environmental Monitoring and Assessment* 85, 277-292.

[15] Noudéhouénou Freedich Madjid HOUESSOU,- *Optimisation de l'effort d'échantillonnage dans le temps et dans l'espace* - mémoire de master 2, 2021.

[16] Rhodes Jonathan R., and Niclas Jonzén. 2011. -*Monitoring temporal trends in spatially structured populations : how should sampling effort be allocated between space and time ?* - Ecography 34 (6) : 1040–48. https://doi.org/10.1111/j.1600-0587.2011.06370.x

[17] Sordello R., Bertheau Y., Coulon A., Jeusset A., Ouédraogo D.Y., Vanpeene S., Vargac M., Villemey A., Witté I., Reyjol Y., Touroult J. - *les protocoles expérimentaux en écologie : principaux points clefs.* - UMS PatrNat, CESCO, Irstea, 2019.

[18] Thorson, J.T., Skaug, H., Kristensen, K., Shelton, A.O., Ward, E.J., Harms, J., Benante, J. - *The importance of spatial models for estimating the strength of density dependence.* - Ecology, 2015, https://doi.org/10.1890/14-0739.1

[19] Walvoort D.J.J., Brus D.J., De Gruijter J.J. - *An R package for spatial coverage sampling and random sampling from compact geographical strata by k-means.* - Computers and Geosciences, vol. 36, n° 10, 2010, pp. 1261-1267.

[20] Wikle, C. K., Zammit-Mangion, A. and Cressie, N. - *Spatio-temporal statistics with R.*- 2019, Chapman & Hall/CRC.