



**ED 398 Géosciences, Ressources Naturelles et Environnement**  
**Proposition de sujet de thèse pour la rentrée universitaire 2020-2021**

**1. Modalités d'encadrement**

Unité(s) de recherche au sein de laquelle le doctorat est réalisé : Centre de Géosciences, MINES ParisTech, équipe Géostatistique et BioSP, INRAe

Directeur de l'unité : Lagneau, Vincent, professeur

Directeur(s)\* de thèse (HDR ou équivalent) : **Romary, Thomas, EC, Centre de Géosciences, MINES ParisTech**

Co-directeur\* de thèse (HDR ou équivalent) : **Allard, Denis, DR, BioSP, INRAE**

Co-encadrant (non HDR) : **Desassis, Nicolas, EC, Centre de Géosciences, MINES ParisTech**

**\* Si un seul directeur de thèse est déclaré, il endosse 100% de la responsabilité de la thèse au regard de l'ED. Si 2 directeurs sont déclarés, ils partagent cette responsabilité à 50%. Le taux de responsabilité maximal est fixé à 300%. Les co-encadrants n'entrent pas dans le décompte, quel que soit leur rôle effectif**

2. Titre **Title** : Prédiction spatio-temporelle par équations différentielles partielles stochastiques *Spatio-temporal prediction by stochastic partial differential equations*

3. Adresse courriel du contact scientifique : thomas.romary@mines-paristech.fr

4. Description du projet de thèse [champ libre 1 page max).

Dans un contexte de transition écologique il est crucial de disposer d'outils d'analyse et de prédiction de l'évolution des milieux naturels et des variables climatiques pour la prise de décision et la gestion des mesures d'atténuation ou d'adaptation. De nombreux domaines des sciences environnementales cherchent à prédire dans l'espace-temps une variable d'intérêt à partir d'observations en certains points d'un domaine d'étude (spatio-temporel) et de variables explicatives (appelées covariables) connues exhaustivement. Par ailleurs l'explosion informatique et les progrès technologiques des instruments de mesure nous ont fait passer de la gestion de la rareté des données à la gestion de leur abondance. Les méthodes numériques doivent être repensées pour traiter de façon efficace ces jeux de données de très grande taille.

Les statistiques spatio-temporelles se sont longtemps limitées à l'hypothèse d'une structure stationnaire dans l'espace-temps. Les enjeux de la thèse sont donc de tirer partie de la richesse des jeux de données actuels, ce qui permet de relaxer cette hypothèse de stationnarité et ainsi améliorer la qualité des prédictions des méthodes géostatistiques. Dans un cadre non-stationnaire, de nombreuses approches ont été développées pour modéliser ces variations spatiales de structure, cf. Fouedjio (2017) ou Schmidt (2020) pour une revue.

L'approche SPDE (Stochastic partial differential equations, Lindgren et al., 2011) permet d'incorporer facilement ces non-stationnarités en faisant varier dans l'espace et dans le temps les coefficients d'un opérateur différentiel. C'est sur cette approche SPDE que nous proposons de nous appuyer pour parvenir à des méthodes de prédiction spatio-temporelles efficaces dans un cadre non-stationnaire. Les travaux engagés dans l'équipe Géostatistique, dont certains en collaboration avec BioSP, sont en pointe dans le domaine.

Dans le cadre spatial, des avancées mathématiques et algorithmiques (Carrizo et al, 2018 ; Pereira & Desassis, 2018 ; Pereira & Desassis, 2019) majeures ont été accomplies, permettant de traiter de façon efficace des jeux de données de très grande taille. Par ailleurs la thèse de Ricardo Carrizo-Vergara (2018) a permis de définir de nouveaux modèles spatio-temporels dans ce cadre, incorporant les processus physiques liés aux phénomènes étudiés (convection, diffusion,...). Nous savons actuellement simuler ces modèles mais les problèmes liés à l'inférence et au conditionnement par les données observées restent entiers. L'objectif de ce projet de thèse est donc de proposer des méthodes efficaces pour l'inférence et la prédiction dans un cadre spatio-temporel, non stationnaire, basé sur l'approche SPDE.

Ce type d'approche peut s'appliquer dans un grand nombre de domaines des géosciences, par exemple le climat, la prédiction de la qualité de l'air en milieu urbain ou celle de l'eau au sein des nappes phréatiques, la quantification de la ressource hydrique, le suivi de données de sol, notamment l'évolution des stocks de carbone organique dans le sol.

### Références

- Carrizo-Vergara, R., Allard, D., & Desassis, N. (2018). A general framework for SPDE-based stationary random fields. arXiv preprint arXiv:1806.04999.
- Carrizo-Vergara, R., (2018) Development of geostatistical models using stochastic partial differential equations, PhD thesis, December 2018. <http://www.theses.fr/2018PSLEM062>
- Fouedjio, F. (2017) Second-order non-stationary modeling approaches for univariate geostatistical data. *Stochastic Environmental Research and Risk Assessment*, 31(8). 1887-1906.
- Fuglstad G.A. , Lindgren F., Simpson D., and Rue H. (2015) Exploring a new class of non-stationary spatial gaussian random fields with varying local anisotropy. *Statistica Sinica*, 115-133.
- Lindgren F., Rue H., Lindström J. (2011) An explicit link between Gaussian fields and Gaussian Markov random fields : the stochastic partial differential equation approach. *Journal of the Royal Statistical Society : Series B*, 73(4) :423-498
- Pereira, M., Desassis, N. (2018). Finite element approximation of non-Markovian random fields. arXiv preprint arXiv:1811.03004.
- Pereira, M., Desassis, N. (2019). Efficient simulation of Gaussian Markov random fields by Chebyshev polynomial approximation. *Spatial Statistics*, 31, 100359.
- Mike Pereira. Champs aléatoires généralisés définis sur des variétés riemanniennes : théorie et pratique. PhD thesis, November 2019. <http://www.theses.fr/s172855>.
- Schmidt, A. M., & Guttorp, P. (2020). Flexible spatial covariance functions. *Spatial Statistics*, <https://doi.org/10.1016/i.spasta.2020.100416>.

## 5. Compétences et connaissances requises [champ libre 1/2 page max.]

De bonnes connaissances en probabilités, statistiques et analyse numérique sont requises, ainsi qu'un intérêt prononcé pour les applications en sciences de l'environnement. Un goût pour la programmation numérique est demandé, et notamment une bonne connaissance des langages C, R et/ou python. La maîtrise de la langue anglaise est aussi nécessaire.

Une lettre de motivation, un descriptif des travaux de stage de Master 2, les résultats d'examen de Master 1 et 2, ainsi que deux lettres de recommandation ou deux référents constitueront les pièces à apporter au dossier de candidature.

## 6. Conditions matérielles de réalisation du projet de recherche

Financement spécifiques obtenus pour le projet : Non

Financement des missions nécessaires pour la réalisation du projet : Non

Accès à des bases de données spécifiques : Non

Accès à des ressources documentaires spécifiques : Non

Accès à des plateformes : Non

Accès à des grands instruments : Non

Autres : **[Champ libre]**

**7. Précisions sur les objectifs de valorisation** des travaux issus du projet de recherche : **[champ libre]**

Cette thèse doit déboucher sur un ensemble de résultats mathématiques et de méthodes sous forme algorithmique. Les premiers seront publiés dans les revues scientifiques et présentés dans des conférences internationales du domaine. Les algorithmes seront codés et diffusés à travers des librairies et packages ouverts.

*Exemples : projet de brevet, types de revues/colloques envisagés/réalisés, actions de vulgarisation scientifique envisagés/réalisés, etc.*

**Visa de la Direction de l'Unité**

**Commentaires éventuels :**