

Lucia CLAROTTO

## Prédiction spatio-temporelle par équations différentielles partielles stochastiques

### *Spatio-temporal prediction by stochastic partial differential equations*

**Résumé du projet de thèse:** 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. Pour cela on dispose de deux approches complémentaires : Les approches classiques de régression (par exemple Random Forest) et la géostatistique et les statistiques spatio-temporelles modélisent la structure de dépendance spatio-temporelle de la variable d'intérêt (ou d'un résidu de celle-ci après avoir tenu compte de la tendance pilotée par les covariables) afin de prédire la variable d'intérêt en un lieu et/ou un temps où elle n'a pas été mesurée ; en outre elle fournit une quantification de l'incertitude liée à cette prédiction. Cependant, 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.

**Thesis abstract:** In a context of ecological transition, it is crucial to have tools for analyzing and predicting the evolution of natural environments and climatic variables for decision-making and the management of mitigation or adaptation measures. Many areas of environmental science seek to predict in spacetime a variable of interest from observations at certain points in a field of study (spatio-temporal) and explanatory variables (called covariates) known exhaustively. In addition, the IT explosion and technological progress in measuring instruments have taken us from managing the scarcity of data to managing their abundance. Numerical methods need to be redesigned to efficiently process these very large datasets. Spatio-temporal statistics have long been limited to the hypothesis of a stationary structure in space-time. The challenges of the thesis are therefore to take advantage of the richness of current data sets, which helps to relax this assumption of stationarity and thus improve the quality of the predictions of geostatistical methods. In a non-stationary framework, many approaches have been developed to model the spatial variations in structure, cf. Fouedjio (2017) or Schmidt (2020) for a review. The SPDE approach (Stochastic partial differential equations, Lindgren et al., 2011) allows these non-stationary factors to be easily incorporated by varying the coefficients of a differential operator in space and over time. It is on this SPDE approach that we propose to rely to arrive at effective spatio-temporal prediction methods in a non-stationary framework. The work undertaken in the Geostatistics team, some of which in collaboration with BioSP, is at the forefront in the field. In the spatial framework, major mathematical and algorithmic advances (Carrizo et al, 2018; Pereira & Desassis, 2018; Pereira & Desassis, 2019) have been made, making it possible to efficiently process very large data sets. In addition, Ricardo Carrizo-Vergara's thesis (2018) made it possible to define new spatio-temporal models in this context, incorporating the physical processes linked to the studied phenomena (convection, diffusion, ...). We currently know how to simulate these models but the problems related to inference and conditioning by the observed data remain intact. The objective of this thesis project is therefore to propose efficient methods for inference and prediction in a spatio-temporal, non-stationary framework, based on the SPDE approach. This type of approach can be applied in a large number of domains of the geosciences, for example the climate, the prediction of air quality in urban areas or that of groundwater, the quantification of water resources, monitoring of soil data, in particular the evolution of organic carbon stocks in the soil.