



HAL
open science

Influence of contextual elements on the spatial modelling of atmospheric cadmium accumulated by mosses

Jérémy Lamouroux, Caroline Meyer, Sébastien Leblond, Isabelle Albert

► To cite this version:

Jérémy Lamouroux, Caroline Meyer, Sébastien Leblond, Isabelle Albert. Influence of contextual elements on the spatial modelling of atmospheric cadmium accumulated by mosses. *Spatial Statistics 2023: Climate and the Environment*, Jul 2023, Boulder (Colorado), United States. hal-04241412

HAL Id: hal-04241412

<https://hal.science/hal-04241412>

Submitted on 13 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Influence of contextual elements on the spatial modelling of atmospheric cadmium accumulated by mosses

Jérémy Lamouroux, Caroline Meyer, Sébastien Leblond, Isabelle Albert

Context

Since the 1960s, terrestrial mosses have been used for monitoring studies of trace element deposits. The **BRAMM** network monitors these elements in rural areas, under tree cover, using monospecific moss samples evenly spread across the territory. Investigating metal accumulation in mosses with contextual elements provides valuable insights for policymakers.

This study focuses on regional **cadmium** (observed $cd \in [0.03; 1.16] \mu\text{g/g}$) deposition and the production of spatial distribution maps. We compare cd accumulation with several contextual elements in **445** French forested sites separated according to bioclimatic zones. Contextual elements used are the concentration of the cd in the atmosphere and cd deposition modelled by EMEP, type of moss sampling, type of forest cover, distance to road, and land use defined by Corine Land Cover. This gives us a total of 30 covariates.

We use **linear models** if no spatial effects are present in the area. Otherwise, we use **kriging models** or enhance the linear model by incorporating a spatial random effect using the Stochastic Partial Differential Equations (**SPDE**) method, which considers spatial correlation. Our results show that the significant contextual elements differ according to the bioclimatic zones. The principal objective is to produce spatial models in which the coefficients linking moss concentration and contextual elements will be calculated according to defined geographical areas.

Method: constrained linear model

Log-log model

We consider an **AIC** selection on $X_{i-1,30}$ with covariates chosen by the experts to avoid correlations between them. The backward-forward selection gives us conserved covariates $X_{i-1,n}$ with n the number of covariates selected in each biogeographical zone or all over France.

For each covariate:

- α is the intercept
- β_i is the estimated coefficient of covariates X_i
- ε_i is the white noise iid

We define the **log-log model** as:

$$\log(\log(1.5/Y)) = \alpha + X_i\beta + \varepsilon_i$$

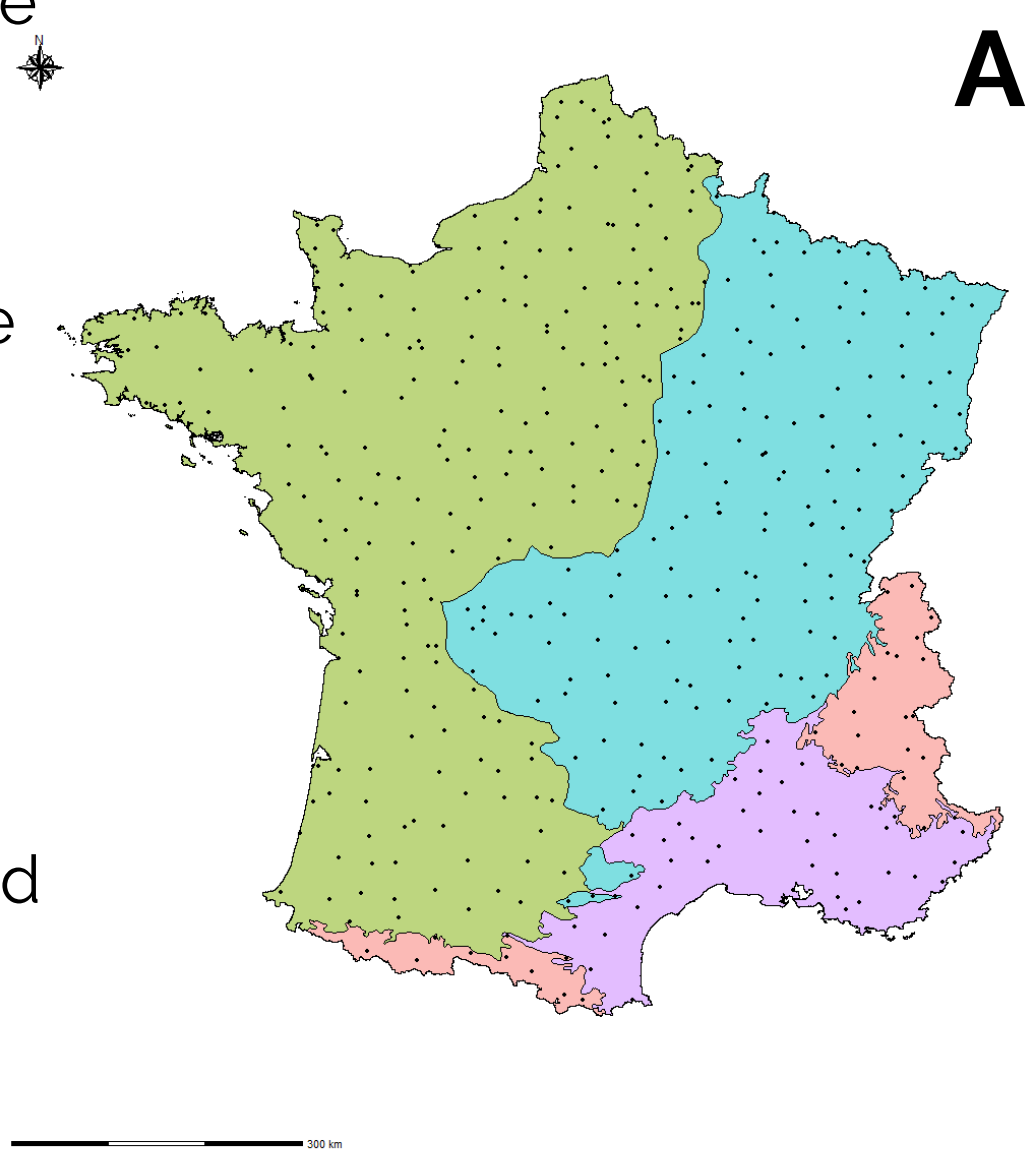
This model was chosen to restrict maximum predictions to 1.5.

Data used and covariates selected

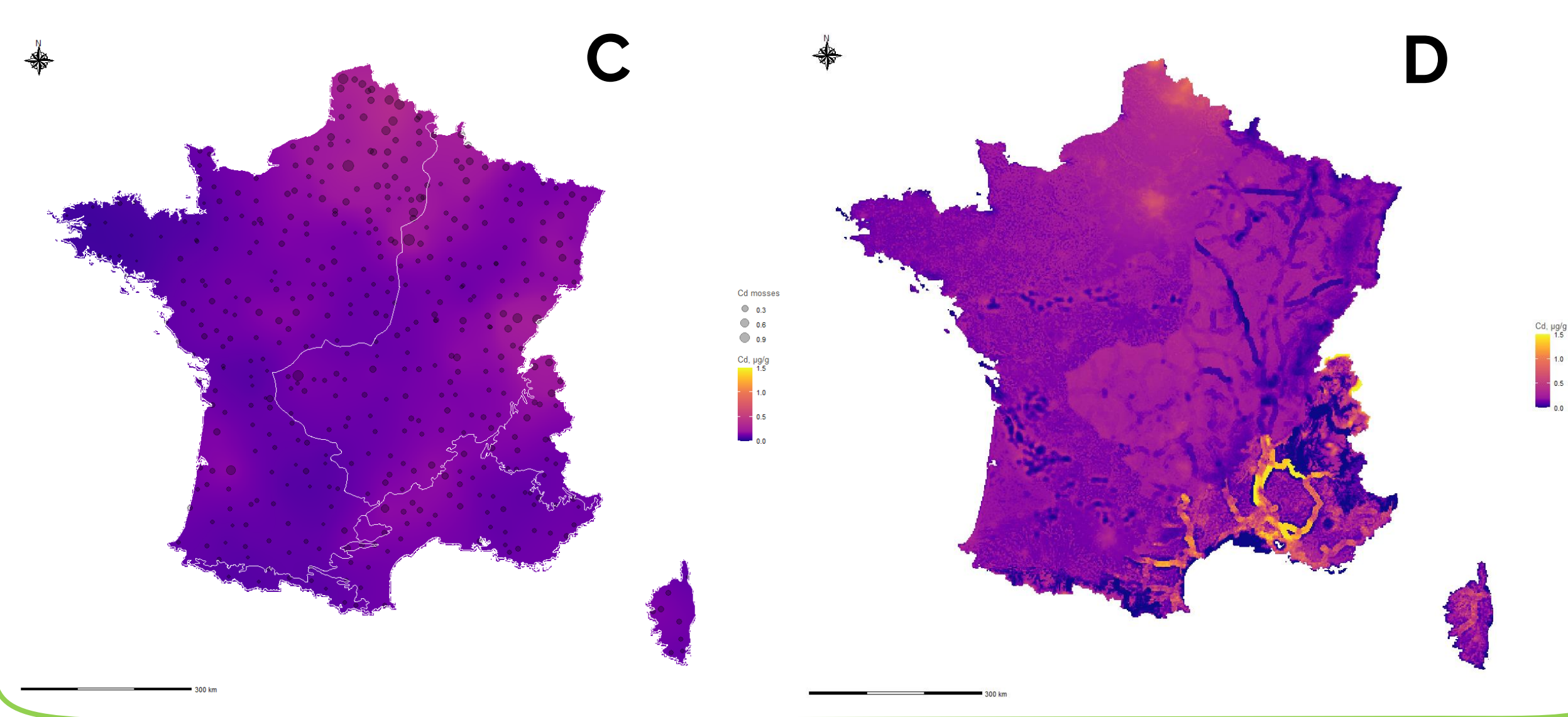
- 445 samples of mosses for **France** and 6 covariates selected by the AIC criterion, the model (E) contains spatial correlation based on the Moran index

Biogeographical zone (A) and covariates selected by the AIC criterion models (D) have no spatial correlation based on the Moran index:

- 220 mosses for the **atlantic** zone and 7 covariates
- 147 mosses for the **continental** zone and 10 covariates
- 49 mosses for the **mediterranean** zone and 19 covariates
- 29 mosses for the **mountain** zone and 20 covariates



Ordinary Kriging (C) & Region-specific models (D)



Results

Predictive values obtained through kriging without covariates fall within the range of 0 to 0.35 $\mu\text{g/g}$, whereas the range for the D map lies between 0 and 1.5 $\mu\text{g/g}$ (values constrained by the model). The log-log model demonstrates enhanced predictions for the high cd values (observed $cd \in [0.03; 1.16] \mu\text{g/g}$). Region-specific models provide regionally specific information by selecting covariates that differ from one region to another. However, more observations per region could improve the model performance.

Method: spatial effects

Spatial model

Our model with spatial effects:

$$\log(\log(1.5/Y(s))) = \mu(s) + \omega(s) + \varepsilon(s)$$

With:

- $\mu(s) = X(s)\beta$
- $\varepsilon(s)$ the white noise iid ; $E(\varepsilon(s)) = 0$, $\text{Var}(\varepsilon(s)) = \tau^2$, $\text{cov}(\varepsilon(s), \varepsilon(s')) = 0$
- $\omega(s)$ the second order mean process stationary, i.e., mean-zero process independent of the white noise process, i.e., $E(\omega(s)) = 0$, $\text{Var}(\omega(s)) = \sigma^2$, $\text{cov}(\omega(s), \omega(s')) = \sigma^2\rho(s,s'; \phi)$, where ρ is a valid two-dimensional correlation function

- Letting $\mu(s) = \beta_0 + \beta_1 X(s)$, write $\omega(s) = \beta_0(s)$.
- Then $\beta_0(s)$ can be interpreted as a random spatial adjustment at location s to the overall intercept β_0 .
- We can define $\tilde{\beta}_0(s) = \beta_0 + \beta_0(s)$ as a random intercept process.

SPDE

We use a SPDE approach to add spatiality in our model. He consists to representing a continuous spatial process, i.e., a Gaussian Fields (GF) using a discretely indexed spatial random process, i.e., Gaussian Markov Random Fields (GMRF).

The SPDE is given by:

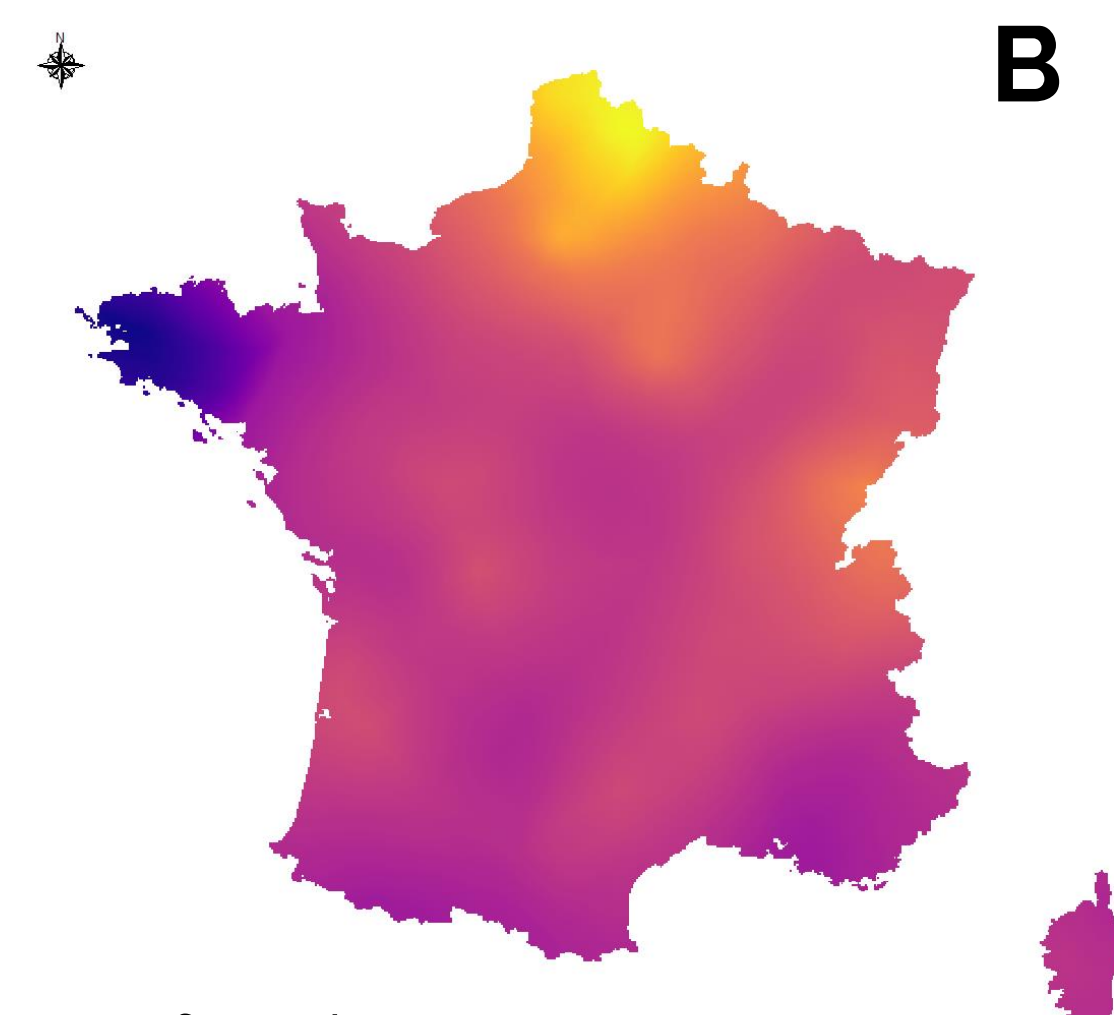
$$(\kappa^2 - \Delta)\alpha / 2(\tau\xi(s)) = \omega(s)$$

With:

- $s \in \mathbb{R}^d$
- Δ is the laplacian
- α controls smoothness
- $K > 0$ is the scale parameter
- T controls the variance
- $\xi(s)$ a GF
- $\omega(s)$ is a gaussian spatial white noise process

The solution of this SPDE is the stationary gaussian

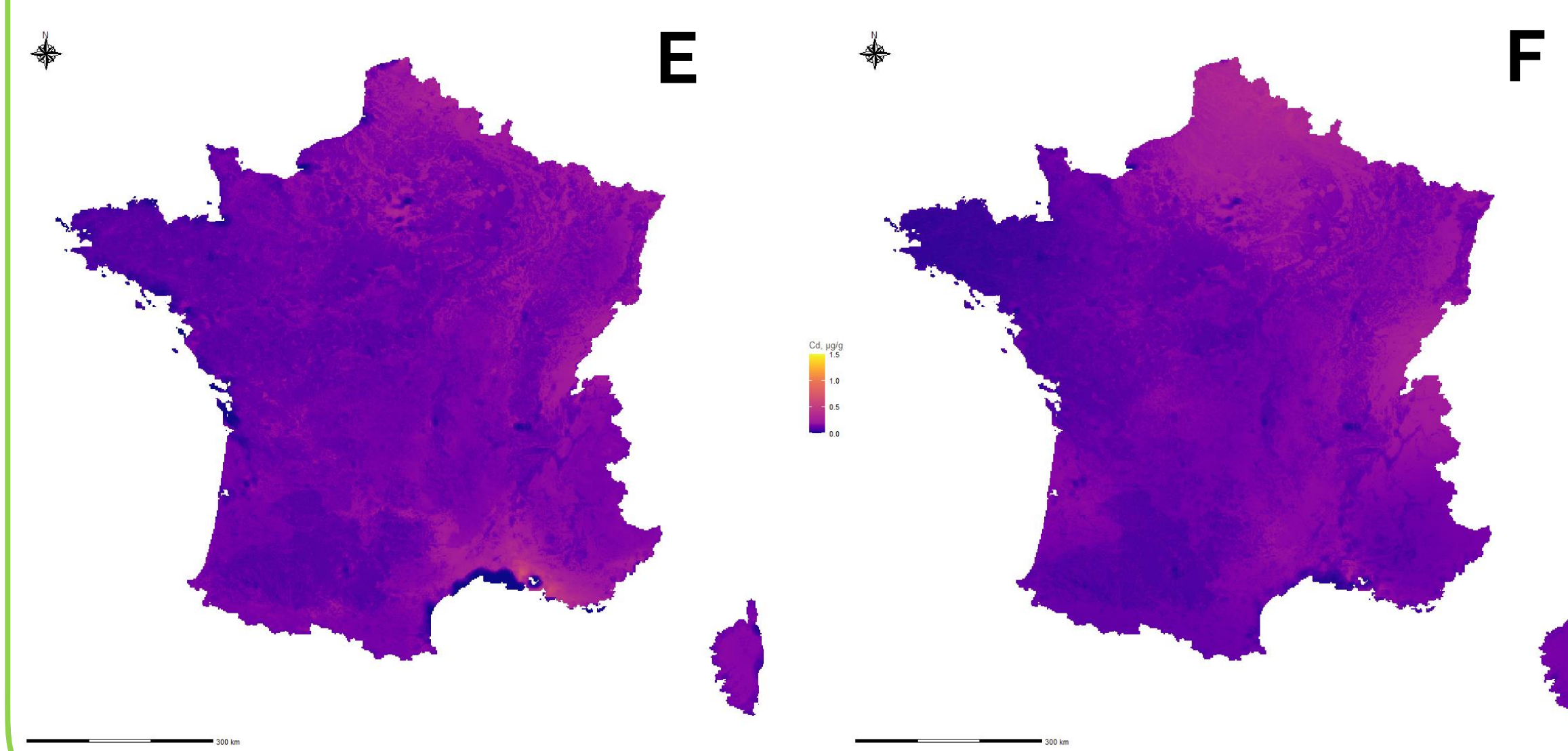
Fields (B) $\xi(s)$ all over France with 6 covariates and a Matérn covariance function.



Model all over France (E) & Prediction of cd concentration from the spatial model (F)

We use **INLA** R package and deviance information criterion (DIC) to predict cd and compare our models with 6 covariates.

- **Figure E**, application of the constrained model all over France without considering the spatial correlation
- **Figure F**, posterior mean of the predicted response variable



Results

The model's predicted values across the whole territory of France (E) exhibit a range of 0 to 0.70 $\mu\text{g/g}$, whereas the spatial model (F) yields values ranging from 0 to 0.45 $\mu\text{g/g}$. The DIC indicates a significant enhancement in results by including a spatial intercept, with values of 218.60 for model E and 150.79 for model F. These findings demonstrate the efficacy of incorporating a spatial intercept to improve the accuracy of our predictions across France.

Conclusion and Perspectives

Our proposed analysis consists of four steps to determine the most accurate variable for explaining the concentration of cd in mosses across France, specifically focusing on the bryologist's perspective. The **D model** allows us to explore the diversity of inputs within different regions of France. We aim to develop a **unified model** (E model) that encompasses the entire country, considering the spatial correlation of mosses (F model). We employed a model incorporating a **spatial intercept** and 6 covariates (selected by the AIC criterion of the linear model) to achieve this F model. The 6 covariates include buffers with a radius of 5km around **urban** and **industrial** areas, 1km around **crops**, 10km around **marine** regions, and **air** and **soil** concentration values from physical models. Our analysis involved comparing the results of this **spatial model** with those of a log-log linear model. The spatial model effectively incorporates **spatial smoothing**, allowing an understanding and interpretation of the spatial variation in cd concentration. On the other hand, the E model lacks spatial smoothing, potentially disregarding significant spatial patterns. In addition to considering spatial correlation in the intercept, the subsequent step was to account for **spatial correlation** in the β coefficients.