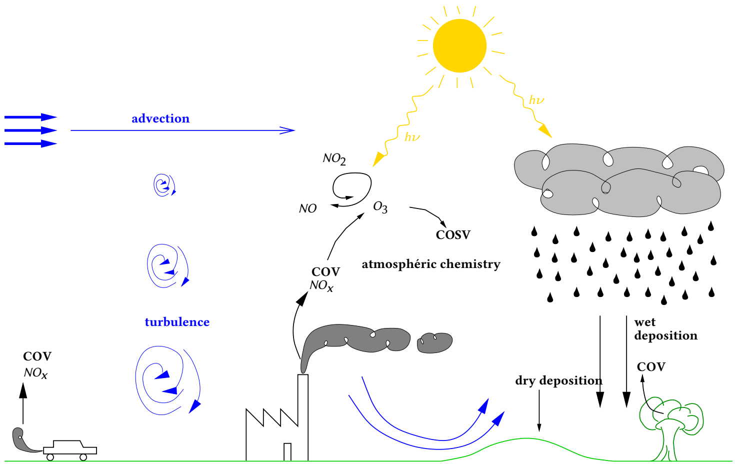
Statistical Modeling and Machine Learning in Meteorology and Oceanography

A geostatistical journey through data and model in air quality

Maxime Beauchamp [1]
Institut Mines-Télécom Atlantique
IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

2020, February 10th

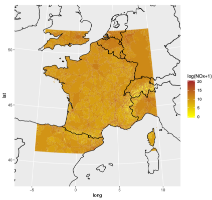
Scientific context

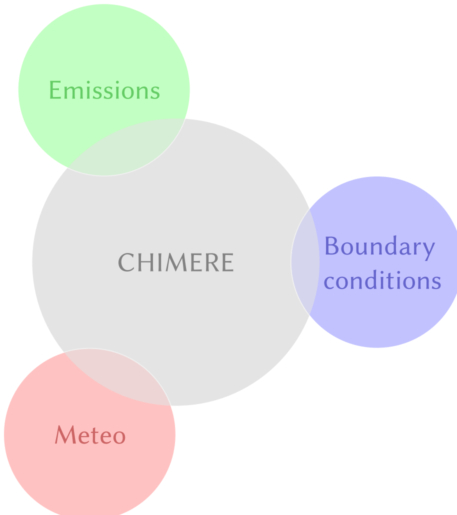


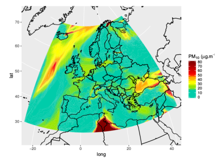
Atmospheric dynamics of the pollutants

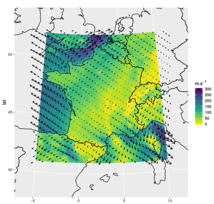
Scientific context I

Implementation of mathematical models to describe the evolution processes of the chemical species (pollutant) in the troposphere

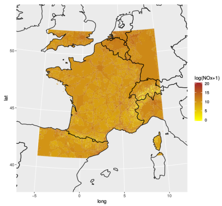


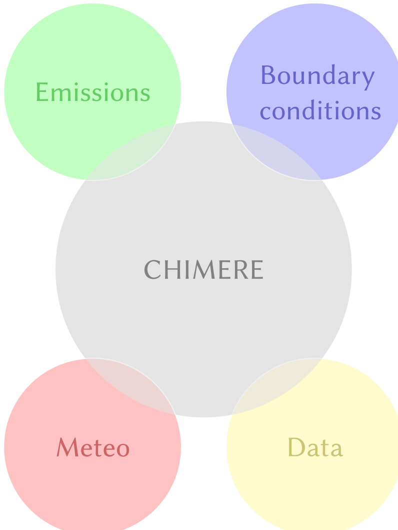


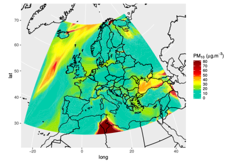


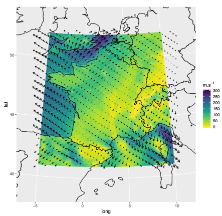


Scientific context II

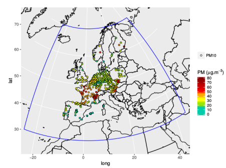








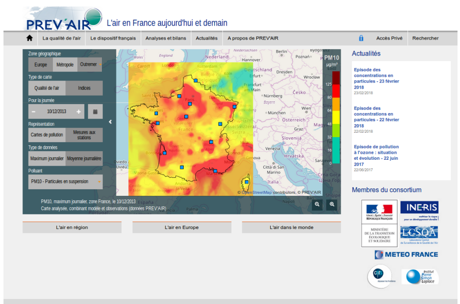
DA, post-processing



PREV'AIR

Operational system for air quality monitoring and forecasting over Europe and France, under the aegis of the Ministry in charge of the environment

- Partners : INERIS, Météo-France, CNRS, IPSL, LCSQA
- Set up in 2003 to deliver daily AQ forecasts and maps on France & Europe
- Based on deterministic chemistry-transport modelling and post-processing using in situ observation data
- During pollution episodes, alert procedures are mainly triggered according to the forecast situation for the previous day (D-1) and next days (D+0, D+1, D+2)

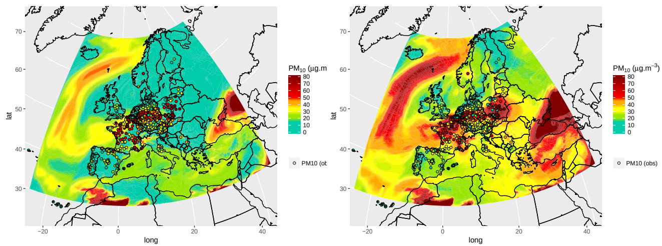


Screenshot of the PREVAIR website `http://www2.prevair.org/`

Two products are delivered by the PREV'AIR system :

I) Analysis (Estimation problem)

Map of the previous day (D-1)
1) Meteorology, Emissions and Boundary conditions are used to run a simulation
2) Monitoring data are collected (France + Europe)
3) Combination of model and data



(a) *CHIMERE simulation*

(b) *Analysis (Kriging with external drift)*

CHIMERE daily simulation and analysis (11th of March 2014)

- Concerning the **analysis problem**, in-situ monitoring observations are generally collected at fixed stations.
- As a consequence, a spatial estimation is sufficient for solving the interpolation
- Different solutions : ...

OI (spatial)

Let note $\mathbf{x}_k \in \mathbb{R}^m$ the locations to interpolate the observations $\mathbf{y}_k \in \mathbb{R}^p$, thus the optimal interpolation (OI) $\mathbf{x}_k^*$ is given by :

$$\mathbf{x}_k^* = \mathbf{\Sigma}_{\mathbf{xy}}\mathbf{\Sigma}_{\mathbf{yy}}^{-1}\mathbf{y}_k$$

with the error covariance matrix :

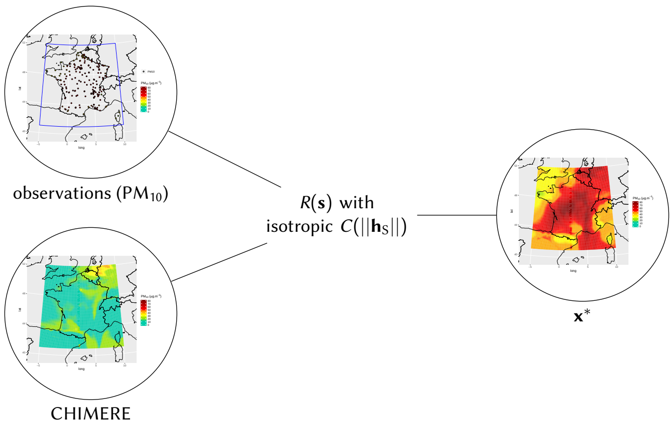
$$\mathbf{P}_k = \mathbf{\Sigma}_{\mathbf{xx}} - \mathbf{\Sigma}_{\mathbf{xy}}\mathbf{\Sigma}_{\mathbf{yy}}^{-1}\mathbf{\Sigma}_{\mathbf{yx}}$$

where :
$\mathbf{\Sigma}_{\mathbf{xy}} = \mathrm{Cov}\{\mathbf{x}_k, \mathbf{y}_k\} \in \mathbb{R}^{m \times p}$
$\mathbf{\Sigma}_{\mathbf{yy}} = \mathrm{Cov}\{\mathbf{y}_k, \mathbf{y}_k\} \in \mathbb{R}^{p \times p}$

OI with covariates / Residual kriging

In the kriging with external drift model (Chiles and Delfiner, 2012), the relation between the covariates φ_l (the model here) and the observations $\mathbf{x}$ is assumed to be linear. At a specific location $\mathbf{s}$

$$\mathbf{x}_\mathbf{s} = \sum_l \beta_l \varphi_l(\mathbf{s}) + R(\mathbf{s})$$



Kriging with external drift

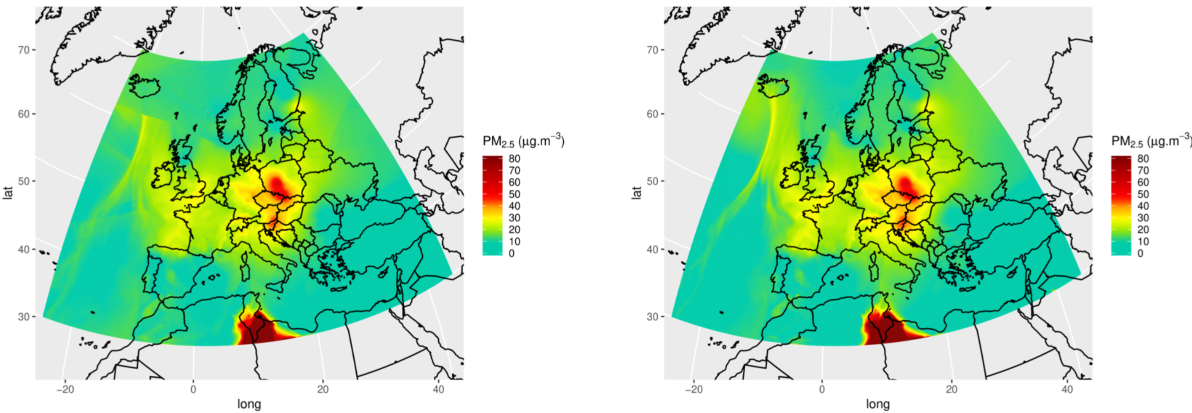
Let note $\mathbf{y}_\alpha = \mathbf{y}(\mathbf{s}_\alpha)$, $\alpha = 1, \cdots, p$. At a particular location $\mathbf{s}_0$ in space, $\mathbf{x}_0^* = \sum_\alpha \lambda_\alpha \mathbf{y}_\alpha$ and the weights λ_α are solution of the linear system (Chiles and Delfiner, 2012) :

$$\begin{cases} \sum_{\alpha=1}^{n} \lambda_\alpha C(\mathbf{s}_\alpha - \mathbf{s}_\beta) + \mu_0 + \sum_{i=1}^{p} \mu_i \varphi_i(\mathbf{s}_\beta) & = C(\mathbf{s}_\beta - \mathbf{s}_0) \quad \forall \beta \\ \sum_{\alpha=1}^{n} \lambda_\alpha & = 1 \\ \sum_{\alpha=1}^{n} \lambda_\alpha \varphi_i(\mathbf{s}_\alpha) & = \varphi_i(\mathbf{s}_0) \quad \forall i \end{cases}$$

$\beta_l(\mathbf{s})$, $i = 0, \cdots, l$ unknown and adjusted within a local neighbourhood

$$\mathbf{x_s} = \sum_l \beta_l(\mathbf{s})\phi_l(\mathbf{s}) + R(\mathbf{s})$$

- Fully bayesian configuration : not only $R(\mathbf{s})$ is stochastic but also the coefficients $\beta_l(\mathbf{x})$.
- Better handling of localisation issues in kriging



Classic Residual vs Bayesian-based residual kriging

In the DA literature, kriging is often presented as a synonym of OI. It is true... but it is also more than that : kriging is a generic term that englobes :

- Optimal interpolation of linear and gaussian quantities
- Optimal estimation of non-linear quantities (Disjunctive kriging, etc.)
- Optimal interpolation of non-gaussian quantities (Poisson, anamorphosis)
- Up/Downscaling of the initial state space considering covariates and support

BLUE (spatial)

In a gaussian context with linear observation operator $\mathbf{H}$, the BLUE is defined by :

$$\mathbf{x}_k^* = \left(\mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k + \mathbf{B}^{-1}\right)^{-1}\left(\mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{y}_k + \mathbf{B}^{-1}\boldsymbol{\mu}_k\right) = \boldsymbol{\mu}_k + \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{H}_k\boldsymbol{\mu}_k\right)$$
$$\mathbf{P}_k = \left(\mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k + \mathbf{B}^{-1}\right)^{-1} = \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}} = \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)\mathbf{B}$$

where the gain matrix $\mathbf{K}_k$ (at time t_k) is defined by :

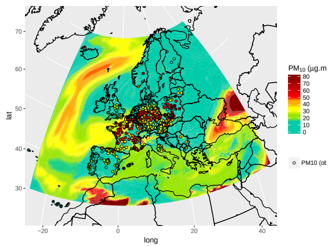
$$\mathbf{K}_k = \mathbf{B}_k\mathbf{H}_k^T\left(\mathbf{R}_k + \mathbf{H}_k\mathbf{B}\mathbf{H}_k^T\right)^{-1}$$

- $\boldsymbol{\mu}_k$ is the expected mean of $\mathbf{x}_k$
- $\mathbf{B} = \mathrm{E}[\mathbf{u}_k\mathbf{u}_k^T]$ is the model error ($\mathbf{u}_k$) covariance matrix
- $\mathbf{R}_k = \mathrm{E}[\mathbf{v}_k\mathbf{v}_k^T]$ is the observation error ($\mathbf{v}_k$) covariance matrix.

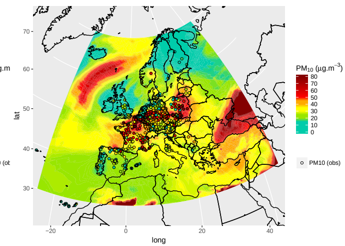
Here, $\mathbf{B}$ is not indexed by k because in a classical OI scheme, it is usually constant and takenb from an historical run.
What is called BLUE in the DA community is also know as multivariate kriging or **cokriging** in a geostatistical setup

II) Forecast (Prediction problem)

Forecast maps of the days D+0, D+1, D+2
1) Meteorology, Emissions and Boundary conditions are used to run a CHIMERE simulation
2) Data from the past
3) Combine model and past data



(a) *CHIMERE simulation*



(b) *Forecast (Kriging with external drift)*

CHIMERE daily simulation and forecast (11th of March 2014)

The **prediction problem** is usually solved by DA techniques (see e.g. Asch et al., 2016)

In AQ, impact(emissions) > impact(initial conditions), thus space-time estimation techniques, i.e. :

- parameter estimation of the deterministic component
- space-time propagation of the stochastic component

are very competitive (MACC project, 2015)

First Idea

Mimic the spatial analysis :

- Because we are dealing with static in-situ observations, provide a statistical model to predict $\mathbf{y}_{k+1}$ from $\mathbf{y}_{1:k}$ and covariates (LGM, GAM, NN, etc.)
- Spatial OI of the statistical forecasts obtained by these in-situ-specific forecasting models

Good to introduce spatial non-stationarity (individual in-situ models) but probably missing space-time correlations.

Space-time residual kriging

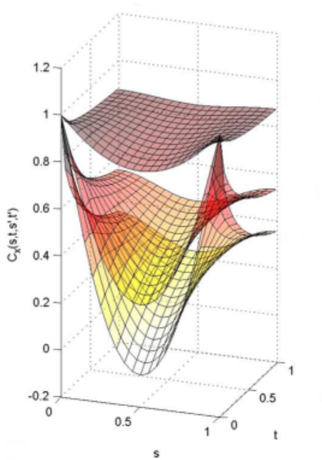
$$\mathbf{x}_k = \mu_k + R_k = \sum \beta_k \varphi_{k,l} + R_k$$

Two options :

- Covariance-based kriging : a simple extension of the spatial kriging, with appropriate space-time covariances
- SPDE-based kriging : a new framework introduced by Lindgren et al. (2011)

Covariance-based kriging

In the tradition framework, a space-time authorized covariance model is given, (see e.g. Gneiting et al., 2007; Porcu et al., 2006; De Iaco et al., 2001)



Examples of space-time covariance

- Major drawbacks : full rank matrices and issues with highly dimensional dataset
- The space-time covariance is generally isotropic in space and symmetric in time

SPDE-based kriging I

The SPDE (spatial) approach is a new formalism which makes it possible to achieve both estimates and simulations. It is constrained by the use of a Matèrn covariance model (which is very general)

$$C(\mathbf{h}) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{h}{a} \right)^{\nu} \mathcal{K}_\nu \left(\sqrt{2\nu} \frac{h}{a} \right)$$

where :

- $\mathcal{K}_\nu$ is the 2nd order modified Bessel function
- ν is the covariance regularity parameter.

Let note that :

- for $\nu = 1/2$, the Matèrn covariance becomes exponential
- $\nu \to +\infty$, the Matèrn covariance becomes gaussian

According to Whittle (1954), if the process $\mathbf{x}_s$, $\mathbf{s} \in \Omega$ is a Gaussian random field (GRF) with Matèrn covariance, then it is solution of the stochastic SPDE :

$$(\kappa^2 - \Delta)^{\alpha/2} \tau \mathbf{x}(s) = \mathcal{W}(s)$$

with :

- $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial s_i^2}$ the Laplacian operator
- $\mathcal{W}(\mathbf{s})$ a standard gaussian white noise
- $\kappa = 1/a$
- $\alpha = \nu + d/2$
- $\tau = \frac{\sigma \Gamma(\nu + d/2)^{1/2} (4\pi)^{d/4} \kappa^{\nu}}{\Gamma(\nu)^{1/2}}$

Results I

On a 2D regular grid, discretizing the SPDE by a FD scheme gives (for $\alpha = 2$) :

$$\kappa^2 \mathbf{x}_{i,j} - \frac{1}{dx}\left(\frac{\mathbf{x}_{i+1,j} - \mathbf{x}_{i,j}}{dx} - \frac{\mathbf{x}_{i,j} - \mathbf{x}_{i-1,j}}{dx}\right) - \frac{1}{dy}\left(\frac{\mathbf{x}_{i,j+1} - \mathbf{x}_{i,j}}{dy} - \frac{\mathbf{x}_{i,j} - \mathbf{x}_{i,j-1}}{dy}\right) = \tau W_{i,j}$$

which can be rewritten as a *sparse* linear system :

$$A\mathbf{x} = \mathbf{W}$$

By denoting $\boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{x})$:

$$A\Sigma A^t = \tau^2 I$$

The *précision* matrix $\mathbf{Q} = \boldsymbol{\Sigma}^{-1}$ is then symmetric and sparse :

$$\mathbf{Q} = \boldsymbol{\Sigma}^{-1} = \frac{1}{\tau^2} A^{\mathsf{T}} A$$

A bridge to kriging I

Let note

$$Z = \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix}$$

The covariance matrix of Z is :

$$\mathrm{Cov}(Z) = \Sigma = \begin{pmatrix} \mathbf{\Sigma}_{YY} & \mathbf{\Sigma}_{YX} \\ \mathbf{\Sigma}_{XY} & \mathbf{\Sigma}_{XX} \end{pmatrix}$$

Thus, the *kriging* and its error covariance matrix are given by :

$$\begin{aligned} \mathbf{x}^\star &= \mathbf{\Sigma}_{XY}\mathbf{\Sigma}_{YY}^{-1}\mathbf{y} \\ \mathrm{Cov}(\mathbf{x}^\star - \mathbf{x}) &= \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX} \end{aligned}$$

A bridge to kriging II

Regarding the *precision matrix* :

$$\mathbf{Q} = \mathbf{\Sigma}^{-1} = \begin{pmatrix} \mathbf{Q}_{YY} & \mathbf{Q}_{YX} \\ \mathbf{Q}_{XY} & \mathbf{Q}_{XX} \end{pmatrix}$$

And the *kriging* can also be written as :

$$\begin{aligned} \mathbf{x}^\star &= -Q_{XX}^{-1} Q_{XY} \mathbf{y} \\ \mathrm{Cov}(\mathbf{x}^\star - \mathbf{x}) &= Q_{XX}^{-1} \end{aligned}$$

Concerning the complexity of the linear systems :

- In traditional geostatistics, we solve a system involving $\mathbf{\Sigma}_{YY}$ which has the size of the data set
- In the SPDE approach, we solve a system involving $\mathbf{Q}_{XX}$ which has the size of the number of target points (state space). However, $\mathbf{\Sigma}_{YY}$ is dense while Q_{XX} is sparse (algorithms adapted to sparse matrices of complexity $\mathcal{O}(n^{3/2})$, while a general Cholesky is of $\mathcal{O}(n^3)$.

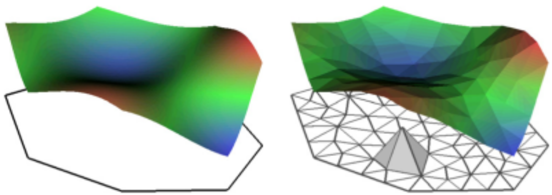
Cas des données irrégulières I

If irregular data $\mathbf{y}$, the finite difference method can no longer be used : use of Finite element method (FEM).
An intermediate stage of triangulation of the domain is then required where data and target points must be located on the vertices of the triangulation.

- The random function $\mathbf{x}$ is then approached by :

$$\mathbf{x}(\mathbf{s}) = \sum_{k=1}^{K} \lambda_k \psi_k(\mathbf{s})$$

where $\psi_k = 1$ at vertex k and linearly decreases to 0 towars neighbouring vertices.



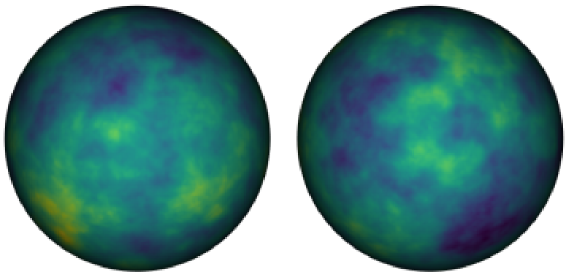
FEM approximation

- The λ_k are random and known at observations locations $\mathbf{y}$
- The precision matrix of Λ is sparse and depends of the triangulation

SPDE perspectives I

For now, the isotropic spatial case was introduces :

$$(\kappa^2 - \Delta)^{\frac{\alpha}{2}} \mathbf{x}(\mathbf{s}) = \tau \mathcal{W}(\mathbf{s})$$



isotropic SPDE-based GRF on the sphere

SPDE perspectives II

A global anisotropy can be added through the SPDE :

$$(\kappa^2 - \nabla \cdot H\nabla)^{\frac{\alpha}{2}} \mathbf{x}(\mathbf{s}) = \tau \mathcal{W}(\mathbf{s})$$

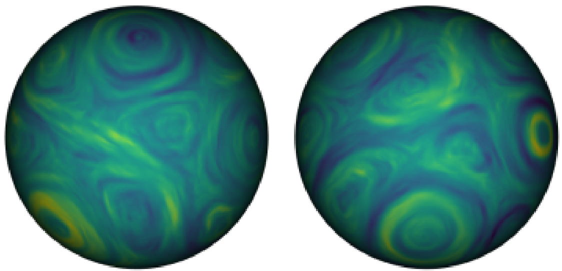
and even better, local anisotropies :

$$(\kappa^2(\mathbf{s}) - \nabla \cdot H(\mathbf{s})\nabla)^{\frac{\alpha}{2}} \mathbf{x}(\mathbf{s}) = \tau(s) \mathcal{W}(\mathbf{s})$$

The complete framework is to estimate space-time processes, non stationary in space and time, with non symmetric and non separable space-time covariances :

$$\left\{ \frac{\partial}{\partial t} + \left(\kappa^2(\mathbf{s}) + \mathbf{m} \cdot \nabla - \nabla \cdot \mathbf{H}(\mathbf{s}) \nabla\right) \right\} \mathbf{x}(\mathbf{s}, t) = \mathcal{W}(\mathbf{s}, t)$$

where **m** is a vector for advection and **H** a diffusion tensor.



anisotropic space-time SPDE-based GRF on the sphere

Application to AQ data I

Model (Cameletti et al., 2012) :

$$\mathbf{x}(\mathbf{s},t) = \underbrace{\beta_0 + \sum \beta_l \varphi_l(\mathbf{s},t)}_{\text{local mean}} + \underbrace{\xi(\mathbf{s},t)}_{\text{latent field}} + \underbrace{\varepsilon(\mathbf{s},t)}_{\text{obs error}}$$

$\varepsilon(\mathbf{s},t) \sim \mathcal{N}(0,\sigma_\varepsilon^2)$ and the latent field is an AR1 process :

$$\xi(\mathbf{s},t) = a\xi(\mathbf{s},t-1) + \omega(\mathbf{s},t)$$

$\omega(\mathbf{s},t) \sim \mathcal{N}(0,\sigma_\omega^2 C(h))$, $C(\mathbf{h})$ a Mátern (spatial) covariance.

Application to AQ data II

A separable space-time covariance is built by approximating the Gaussian field by its Finite Elements representation :

$$\xi(\mathbf{s},t)=\sum_k \psi_l(\mathbf{s},t)\omega_k=\sum_k \psi_i^s(\mathbf{s})\psi_j^t(t)\omega_k$$

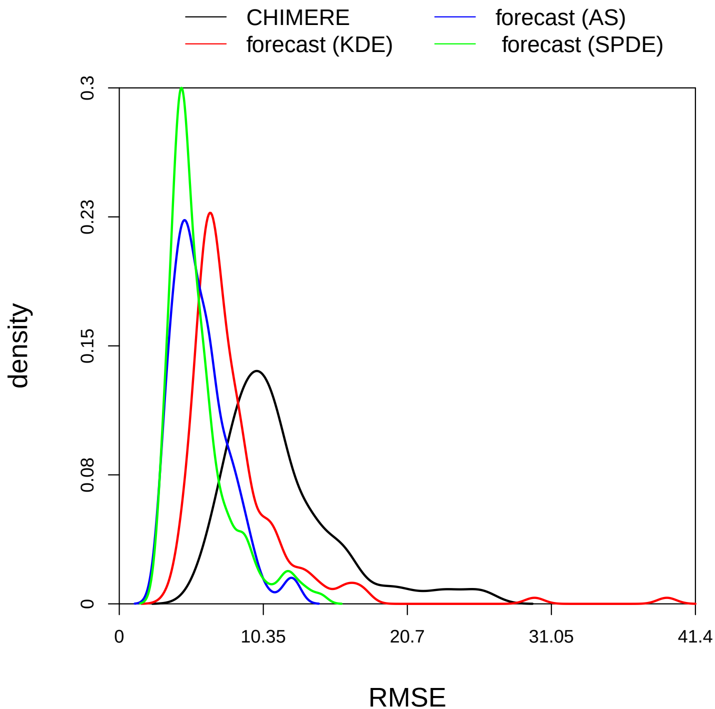
where the basis functions are seen as the product of purely spatial basis functions $\psi_i^s(\mathbf{s})$ and purely temporal basis functions $\psi_j^t(t)$, then the space-time stochastic PDE (Lindgren et al., 2011) defined by :

$$\left\{\frac{\partial}{\partial t}+(\kappa(\mathbf{s})^2-\Delta)^{\alpha/2}\tau(\mathbf{s})\right\}\xi(\mathbf{s},t))=\mathcal{W}(\mathbf{s},t),\quad (\mathbf{s},t)\in\mathcal{D}\times\mathbb{R}$$

generates a precision matrix $\mathbf{Q}$ for the Gaussian weights ω_k so that :

$$\mathbf{Q}=\mathbf{Q_T}\otimes\mathbf{Q_S}$$

$\mathbf{Q_S}$ and $\mathbf{Q_T}$ are respectively the precision matrices of the purely spatial model and the Markovian random walk.



Daily PM_{10} RMSE (French domain) in 2014 for CHIMERE, time+spatial kriging, covariance-based space-time kriging and SPDE-based space-time kriging

Perspectives : Urban scale

- Two scales in AQ : regional/global scale and urban scale
- Urban scale aims at integrating high-resolution emissions (traffic-related, etc.)
- Historically : Few data (less than 10 monitoring sites over the city) or more but with a low temporal resolution (annual mean)
- In the last two years, new framework with mobile sensors : cheap and numerous
- But large and correlated errors ($\mathbf{R}$), provide efficient way to deal with partial and noisy observations

Perspectives : Large scale

Learning from the past

- AnDA provides an interesting framework to avoid running costly CTMs : biogenic and anthropic emissions have clear seasonality that can be used in an analog perspective

Need for metamodeling : sensitivy of CTMs to emission for AQ/climate change scenarii

- SPDE-based simulations, coupled with parameter estimation (Rue et al., 2009)
  Cholesky factorization $\mathbf{Q} = ŁŁ^T$ then solving triangular system $Ł^T\mathbf{Z} = \mathbf{U}$ with a standard gaussian random vector $\mathbf{U}$ with independant components
- Deep-learning NN to learn AQ dynamics conditionnaly to meteorology and emissions

References I

M. Asch, M. Bocquet, and M. Nodet. *Data assimilation : methods, algorithms, and applications.* Fundamentals of Algorithms. SIAM, 2016. URL `https://hal.inria.fr/hal-01402885`.

M. Cameletti, F. Lindgren, D. Simpson, and H. Rue. Spatio-temporal modeling of particulate matter concentration through the spde approach. *AStA Advances in Statistical Analysis*, 97(2) :109–131, 2012. ISSN 1863-818X. doi : 10.1007/s10182-012-0196-3. URL `http://dx.doi.org/10.1007/s10182-012-0196-3`.

J. Chiles and P. Delfiner. *Geostatistics : modeling spatial uncertainty.* Wiley, New-York, second edition, 2012.

S. De Iaco, D. Myers, and D. Posa. Space-time analysis using a general product-sum model. *Statistics & Probability Letters, v. 52, no. 1*, pages 21–28, 2001.

T. Gneiting, M. Genton, and P. Guttorp. *Geostatistical Space-Time Models, Stationarity, Separability, and Full Symmetry*, pages 151–175. C&H/CRC Monographs on Statistics & Applied Probability. Chapman and Hall/CRC, Oct 2007. ISBN 978-1-58488-593-1. doi : 10.1201/9781420011050.ch4. URL `http://dx.doi.org/10.1201/9781420011050.ch4`. 0.

F. Lindgren, H. Rue, and J. Lindström. 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 (Statistical Methodology)*, 73(4) :423–498, 2011. ISSN 1467-9868. doi : 10.1111/j.1467-9868.2011.00777.x. URL `http://dx.doi.org/10.1111/j.1467-9868.2011.00777.x`.

E. Porcu, P. Gregori, and J. Mateu. Nonseparable stationary anisotropic space–time covariance functions. *Stochastic Environmental Research and Risk Assessment*, 21(2) :113–122, 2006. ISSN 1436-3259. doi : 10.1007/s00477-006-0048-3. URL `http://dx.doi.org/10.1007/s00477-006-0048-3`.

References II

H. Rue, S. Martino, and N. Chopin. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 71(2) :319–392, 2009. doi : 10.1111/j.1467-9868.2008.00700.x. URL `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2008.00700.x`.

P. Whittle. On stationary processes in the plane. *Biometrika*, 41(3-4) :434–449, 1954. doi : 10.1093/biomet/41.3-4.434. URL `http://biomet.oxfordjournals.org/content/41/3-4/434.short`.