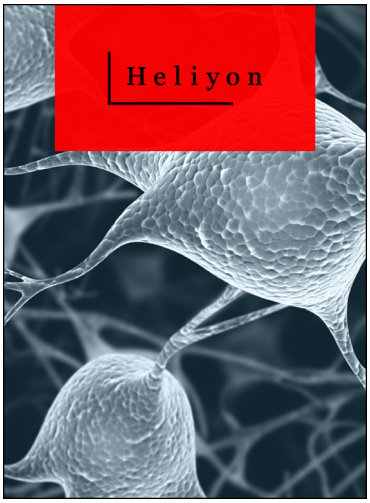
An iterative optimization scheme to accommodate inequality constraints in air quality geostatistical estimation of multivariate PM

Maxime Beauchamp, Bertrand Bessagnet



An iterative optimization scheme to accommodate inequality constraints in air quality geostatistical estimation of multivariate PM

Maxime Beauchamp and Bertrand Bessagnet

PII: S2405-8440(23)04621-2
DOI: https://doi.org/10.1016/j.heliyon.2023.e17413
Reference: HLY e17413

To appear in: *Heliyon*

Received date: 22 August 2022
Revised date: 14 June 2023
Accepted date: 15 June 2023

Please cite this article as: M. Beauchamp and B. Bessagnet, An iterative optimization scheme to accommodate inequality constraints in air quality geostatistical estimation of multivariate PM, *Heliyon*, 9, e17413, doi: https://doi.org/10.1016/j.heliyon.2023.e17413.

An iterative optimization scheme to accommodate inequality constraints in air quality geostatistical estimation of multivariate PM

Maxime Beauchamp[a,*,1], Bertrand Bessagnet[b]

[a]IMT Atlantique Bretagne-Pays de la Loire Campus de Brest Technopôle Brest-Iroise CS 83818 29238 Brest cedex 03 France
[b]European Commission, Joint Research Centre (JRC), Ispra, Italy

ARTICLE INFO

Keywords:
air quality
cokriging
optimization
penalization
multivariate particulate matter
chemistry-transport model

Abstract

The kriging-based estimation of the different types of atmospheric particulate matter (PM) pollutions defined in the air quality regulation raises some operational problems because the (co)kriging equations are obtained by minimizing a linear combination of the estimation variances subject to unbiasedness constraints. As a consequence, the estimation process can result in total PM_{10} concentrations that are less than the $PM_{2.5}$ concentrations which would be physically impossible. In a previous publication, it was shown that a convenient external drift modelling can reduce the number of spatial locations where the inequality constraint is not satisfied, without completely solving the problem. In this work, the formulation of the cokriging system is modified, inspired by previous works focusing on positive kriging. The introduction of additional constraints on the cokriging weights are presented, leading to a unique and optimal solution to the problem of cokriging under inequality constraints between two variables. Some computational and algorithmic details are introduced. An evaluation of the penalized cokriging is provided by using the European PM monitoring sites dataset: some maps and performance scores are given to assess the relevance of our iterative optimization scheme.

1. Introduction

In air quality, the geostatistical estimation is commonly used [9, 19, 20, 22, 30] to produce maps of ambiant air regulatory pollutants [15], including PM_{10} and $PM_{2.5}$, the particles whose diameter are respectively smaller than 10 $\mu g.m^{-3}$ and 2.5 $\mu g.m^{-3}$. If the PM_{10} mapping is a topic already well documented [4, 8], the interest for $PM_{2.5}$ is growing [11, 34, 39] for public health reasons.

Measurements of $PM_{2.5}$ having started later than PM_{10} measurements, their monitoring network is less developed, which can affect the final mapping. In the French PREV'AIR system, see e.g. [33], the analyzed maps of ozone and PM_{10} are built by kriging the observations [26], including simulations of the CTM CHIMERE model [25] as external drift. Applying such a technique to $PM_{2.5}$ does not take into account the inequality between the concentrations of $PM_{2.5}$ and PM_{10}, and can therefore lead to $PM_{2.5}$ estimations greater than PM_{10}. Ad-hoc corrections were already considered, see e.g. [33] to estimate $PM_{2.5}$ as the corresponding PM_{10} times the ratio $PM_{2.5}/PM_{10}$ simulated by CHIMERE. However, this method is not satisfactory given the uncertainties in the model-based simulations, in particular during air pollution episode. For the joint estimation of both pollutants, cokriging is the most common solution used with a linear coregionalization model [38]. Different solutions were already implemented: [35] proposed a Bayesian version of the linear model of coregionalization applied to air quality data (CO, NO, NO_2). Multivariate non-stationarity processes were also addressed by Bayesian formulations, see e.g. [7], with spatial variations of the coefficients in the covariance model [18].

Beauchamp et al. [3] also developed a cokriging model for mapping PM_{10} and $PM_{2.5}$, in order to improve the precision of PM mapping. In this work, the local means of both variables are related by an additive model, but the latter does not ensure to satisfy the physical relationship existing between the two after their joint estimation. Indeed, the $PM_{2.5}$ concentration is by definition less than PM_{10}. At most, the additive model of Beauchamp et al. [3] ensures the drift, a local regression of the observations on the Chemistry Transport Model (CTM) outputs [24], to be physically consistent thanks to the appropriate unbiasedness conditions. Thus, when the correlation between the model and the two observational datasets is good enough, the final estimation is very likely physically consistent. This is no longer the case if the drift poorly approaches the data. Anyway, this cokriging model with additive external drift has demonstrated its ability to better reproduce the PM concentration levels for mapping. It significantly reduces the number of cases in which $PM_{2.5}$ estimations are greater than PM_{10} (from 5 to 0.5%). However, even if the operational solution to deal with the inequality constraint is satisfactory, the mathematical problem associated is not entirely solved.

The way of dealing with inequalities in kriging has already been addressed, but mostly when the only information available is greater or less than a given value [16, 23]. For the estimation itself, Michalak [27] proposed a Gibbs sampler-

*Maxime Beauchamp
maxime.beauchamp@imt-atlantique.fr (M. Beauchamp); bertrand.bessagnet@ec.europa.eu (B. Bessagnet)
ORCID(s): 0000-0003-0605-8935 (M. Beauchamp)

based approach for inequality-constrained geostatistical interpolation, supported by a review on the existing works about this topic: non-negativity of the weights [14], Lagrange multipliers based approaches [28] or Monte-Carlo methodologies [1], etc. Additional equality and inequality constraints may also be used to solve kriging problems, such as negative weights in ordinary and simple kriging [37] or multiple indicator kriging [36].

In this work, a new algorithmic approach is presented to ensure $PM_{2.5}$ estimation be less than PM_{10}, by introducing additional inequality constraints on the cokriging weights, writing the related optimization problem and solving it. The approach relies on Karush-Kuhn-Tucker conditions, inspired by the work of [2] on positive kriging. Additional constraints are formulated on the cokriging weights. When not met, the constraints result in rewriting the cokriging system. It yields an iterative procedure which is used until all the constraints are satisfied to ensure the consistency between both $PM_{2.5}$ and PM_{10} estimation.

2. Methods

In Beauchamp et al. [3], the PM description is made through the simple additive model (1):

$$Z(\mathbf{x}) = Y(\mathbf{x}) + W(\mathbf{x}) \tag{1}$$

with $Y(\mathbf{x}) \leq Z(\mathbf{x})\ \forall \mathbf{x}$.

Here, the joint estimation between $Z(\mathbf{x})$, the PM_{10} concentration and $Y(\mathbf{x})$, the $PM_{2.5}$ concentration, is investigated. Such a geostatistical estimation is known as cokriging [10].

The two components of the sum are supposed to be non-stationary, but can be explained by the use of deterministic covariates $f^i(\mathbf{x})$ and $g^j(\mathbf{x})$, see Eq. (2):

$$\begin{aligned} Y(\mathbf{x}) &= m_Y(\mathbf{x}) + R(\mathbf{x}) \\ &= a_0 + \sum_i^l a_i f^i(\mathbf{x}) + R(\mathbf{x}) \\ W(\mathbf{x}) &= m_W(\mathbf{x}) + T(\mathbf{x}) \\ &= b_0 + \sum_j^p a_j g^j(\mathbf{x}) + T(\mathbf{x}) \end{aligned} \tag{2}$$

$f^i,\ i = 1, \cdots, l$ and $g^j,\ j = 1, \cdots, p$ are the covariates respectively used in the computation of the local means $m_Y(\mathbf{x})$ and $m_W(\mathbf{x})$. l and p are the number of covariates for the regression of Y on the f^i and of W on the g^j. Last, R and T are residuals assumed to be second-order stationary random functions with zero mean. This framework can be used for any similar modelling, and even simplified if not using any covariates.

Let denote $Y^{\mathrm{CK}}(\mathbf{x}_0)$ and $Z^{\mathrm{CK}}(\mathbf{x}_0)$ the cokriging-based estimation of $PM_{2.5}$ and PM_{10} at location $\mathbf{x}_0$. For sake of simplicity, Y and Z respectively refer to variables with index 1 and 2 throughout the paper. We aim at producing a joint estimation such that $Y^{\mathrm{CK}}(\mathbf{x}_0)$ is less than $Z^{\mathrm{CK}}(\mathbf{x}_0)$. The additive relationship between Z and Y implies that direct and cross covariances are linked through the following set of equations:

$$\begin{cases} C_Z(\mathbf{h}) &= C_R(\mathbf{h}) \\ C_{Z,Y}(\mathbf{h}) &= C_R(\mathbf{h}) + C_{R,T}(\mathbf{h}) \\ C_{Y,Z}(\mathbf{h}) &= C_R(\mathbf{h}) + C_{T,R}(\mathbf{h}) \\ C_Y(\mathbf{h}) &= C_R(\mathbf{h}) + C_{R,T}(\mathbf{h}) \\ & \quad + C_{R,T}(-\mathbf{h}) + C_T(\mathbf{h}) \end{cases}$$

The same conditions can also be written using variograms:

$$\begin{cases} \gamma_Z(\mathbf{x}, \mathbf{x}+\mathbf{h}) &= C_R(0) - C_R(\mathbf{h}) \\ \gamma_{Y,Z}(\mathbf{x}, \mathbf{x}+\mathbf{h}) &= \gamma_R(\mathbf{h}) + \gamma_{R,T}(\mathbf{h}) \\ \gamma_Y(\mathbf{x}, \mathbf{x}+\mathbf{h}) &= \gamma_R(\mathbf{h}) + \gamma_T(\mathbf{h}) \\ & \quad + 2\gamma_{R,T}(\mathbf{h}) \end{cases}$$

$C(.)$ denotes a covariance and $\gamma(.)$ a variogram, see e.g. Cressie and Wikl [13]. In Beauchamp et al. [3], it is described why the use of variograms instead of covariances is preferable and well suited to the data and their related hypothesis.

$Z^{\mathrm{CK}}(\mathbf{x}_0)$ is obtained by solving the cokriging system (CK), see Eq. (3) and Sect. 2.5 in [3]:

$$Z^{\mathrm{CK}}(\mathbf{x}_0) = \sum_{\beta=1}^{M} \nu'_\beta Z(\mathbf{x}_\beta) + \sum_{\alpha=1}^{N} \lambda'_\alpha Y(\mathbf{x}_\alpha) \tag{3}$$

where $\alpha = 1, \cdots, N$ and $\beta = 1, \cdots, M$ are the indices for the observations $Y(\mathbf{x}_\alpha)$ and $Z(\mathbf{x}_\beta)$. $\nu'_\beta = \nu'(\mathbf{x}_\beta)$ and $\lambda'_\alpha = \lambda'(\mathbf{x}_\alpha)$ are the cokriging weights of the observations $Z(\mathbf{x}_\beta)$ and $Y(\mathbf{x}_\alpha)$. Modifying a bit the cokriging system notations of Beauchamp et al. [3] to ease the link with this work, see Eq. (4), the cokriging system of $Z^{\mathrm{CK}}(\mathbf{x}_0)$ becomes:

$$\begin{bmatrix} \mathbf{K}_{22} & \mathbf{K}_{21} & 1 & \mathbf{g}_2^j & 0 & \mathbf{f}_2^i \\ \mathbf{K}_{12} & \mathbf{K}_{11} & 0 & 0 & 1 & \mathbf{f}_1^i \\ 1 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{g}_2^j & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \mathbf{f}_2^i & \mathbf{f}_1^i & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\nu}' \\ \boldsymbol{\lambda}' \\ \mu_0 \\ \boldsymbol{\mu_j} \\ \upsilon_0 \\ \boldsymbol{\upsilon_i} \end{bmatrix} = \begin{bmatrix} \mathbf{K}_{20} \\ \mathbf{K}_{120} \\ 1 \\ g_0^j \\ 0 \\ f_0^i \end{bmatrix} \tag{4}$$

where:
$\mathbf{K}_{11} = \{C_Y(\mathbf{x}_\alpha - \mathbf{x}_{\alpha'})\}$.
$\mathbf{K}_{22} = \{C_Z(\mathbf{x}_\beta - \mathbf{x}_{\beta'})\}$.
$\mathbf{K}_{12} = \{C_{Y,Z}(\mathbf{x}_\alpha - \mathbf{x}_\beta)\}$.
$\mathbf{K}_{20} = \{C_Z(\mathbf{x}_\beta - \mathbf{x}_0)\}$.

$\mathbf{K}_{120} = \{C_{Y,Z}(\mathbf{x}_\alpha - \mathbf{x}_0)\}$.
$\boldsymbol{\nu}' = (\nu'_\beta)$ and $\boldsymbol{\lambda}' = (\lambda'_\alpha)$ are the cokriging weights of the observations $Z(\mathbf{x}_\beta)$ and $Y(\mathbf{x}_\alpha)$. $\mathbf{g}_2^j$ denotes the vector of the jth covariate g^j at locations $\mathbf{x}_\beta$ $\{g^j(\mathbf{x}_\beta)\}$. $\mathbf{f}_2^i$ denotes the vector of the ith covariate f^i at locations $\mathbf{x}_\beta$ $\{f^i(\mathbf{x}_\beta)\}$. $\mathbf{f}_1^i$ denotes the vector of the ith covariate f^i at locations $\mathbf{x}_\alpha$ $\{f^i(\mathbf{x}_\alpha)\}$. g_0^j and f_0^i respectively denote the jth and ith covariates g^j and f^i at the target location $\mathbf{x}_0$. Last, μ_0, $\boldsymbol{\mu_j}$, υ_0, $\boldsymbol{\upsilon_i}$ are the Lagrange multipliers for the unbiasedness conditions, see Eq.(11a) to Eq.(11d) in Beauchamp et al. [3].

For sake of generality, this system is given with covariances but can be easily written with variograms, as this tool was used in Beauchamp et al. [3] and again in this study. As denoted by Isaaks and Srivastava [21], this type of cokriging estimation may lead to negative values and the effect of the secondary variable on the estimation is weak depending on the spatial sampling of the two variables, but it is not relevant here because there is generally more PM_{10} data available than $PM_{2.5}$ and the deterministic covariate almost ensures the positivity of the estimation, see again Beauchamp et al. [3].

In a more convenient synthetic notations:

$$\begin{aligned} \mathbf{KP} + \mathbf{AM} &= \mathbf{K_0} \\ \mathbf{A}'\mathbf{P} &= \mathbf{F_0} \end{aligned} \tag{5}$$

where :

$$\mathbf{K} = \begin{bmatrix} \mathbf{K_{11}} & \mathbf{K_{12}} \\ \mathbf{K_{21}} & \mathbf{K_{22}} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1 & \mathbf{g}_2^j & 0 & \mathbf{f}_2^i \\ 0 & 0 & 1 & \mathbf{f}_1^i \end{bmatrix},$$

$$\mathbf{P} = \begin{bmatrix} \nu'_\beta \\ \lambda'_\alpha \end{bmatrix}, \quad \mathbf{M} = \begin{bmatrix} \boldsymbol{\mu_0} \\ \boldsymbol{\mu_j} \\ \boldsymbol{\upsilon_0} \\ \boldsymbol{\upsilon_i} \end{bmatrix}$$

$$\mathbf{K_0} = \begin{bmatrix} \mathbf{K}_{20} \\ \mathbf{K}_{120} \end{bmatrix}, \quad \mathbf{F_0} = \begin{bmatrix} 1 \\ g_0^j \\ 0 \\ f_0^i \end{bmatrix}$$

Let suppose now that $Z^{\mathrm{CK}}(\mathbf{x}_0)$ from Eq. (5) is already computed. The estimation $Y^{\mathrm{CK}}(\mathbf{x}_0)$ writes:

$$Y^{\mathrm{CK}}(\mathbf{x}_0) = \sum_{\alpha=1}^{N} \lambda_\alpha Y(\mathbf{x}_\alpha) + \sum_{\beta=1}^{M} \nu_\beta Z(\mathbf{x}_\beta)$$

where the weights λ_α of $Y(\mathbf{x}_\alpha)$ and the weights ν_β of $Z(\mathbf{x}_\beta)$ are found by solving a cokriging system similar to (4), wherein the order of the variables is simply modified accordingly in matrix $\mathbf{K}$ and vectors $\mathbf{P}$, $\mathbf{M}$ and $\mathbf{K_0}$. The unbiasedness conditions also modify the contents of matrix $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & \mathbf{f}_1^i \\ 0 & \mathbf{g}_2^j & 1 & \mathbf{f}_2^i \end{bmatrix}, \mathbf{F_0} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ f_0^i \end{bmatrix}$$

2.1. Additional constraints on the weights

$Y^{\mathrm{CK}}(\mathbf{x}_0)$ has to be less than the prior estimation $Z^{\mathrm{CK}}(\mathbf{x}_0)$ of the first variable, i.e.:

$$\begin{aligned} Y^{\mathrm{CK}}(\mathbf{x}_0) &\leq Z^{\mathrm{CK}}(\mathbf{x}_0) \\ &\leq \sum_{\beta=1}^{M} \nu'_\beta Z(\mathbf{x}_\beta) + \sum_{\alpha=1}^{N} \lambda'_\alpha Y(\mathbf{x}_\alpha) \end{aligned} \tag{6}$$

Let precise that cutting the estimation of Y^{CK} as the minimum between $(Y^{\mathrm{CK}}, Z^{\mathrm{CK}})$ is not a good solution because even if it can lead to satisfying results, both in terms of mapping and obviously being consistent with the physical constraints, it does not enable to compute the corresponding cokriging standard deviation.

Replacing the cokriging of Z by its mathematical expression given in Eq.3 leads to:

$$\sum_{\alpha=1}^{N} \lambda_\alpha Y(\mathbf{x}_\alpha) \leq \sum_{\alpha=1}^{N} \lambda'_\alpha Y(\mathbf{x}_\alpha) + \sum_{\beta=1}^{M} (\nu'_\beta - \nu_\beta) Z(\mathbf{x}_\beta)$$

Revisiting this inequality, we make the choice of evenly decomposing $\sum_{\beta=1}^{M} (\nu'_\beta - \nu_\beta) Z(\mathbf{x}_\beta) = \sum_{\alpha=1}^{N} \tilde{\lambda}'_\alpha$ as the sum of N terms where:

$$\tilde{\lambda}'_\alpha = \left[\sum_{\beta=1}^{M} (\nu'_\beta - \nu_\beta) Z(\mathbf{x}_\beta) \right] \frac{l(\mathbf{x}_\alpha)}{Y(\mathbf{x}_\alpha)}, \quad \alpha = 1, \cdots, N$$

with $\sum_{\alpha=1}^{N} l(\mathbf{x}_\alpha) = 1$ and $Y(\mathbf{x}_\alpha) > 0 \ \forall \alpha$. In this work, we consider a uniform repartition scheme, then $l(\mathbf{x}_\alpha) = 1/N$ $\forall \mathbf{x}_\alpha$, but any other scheme could be experimented. The additional weights $\tilde{\lambda}'_\alpha$ are then distributed over the N weights λ'_α:

$$\sum_{\alpha=1}^{N} \lambda_\alpha Y(\mathbf{x}_\alpha) \leq \sum_{\alpha=1}^{N} \left[\lambda'_\alpha + \tilde{\lambda}'_\alpha \right] Y(\mathbf{x}_\alpha)$$

Thus, at ν_β fixed, the condition (7):

$$\lambda_\alpha \leq \lambda'_\alpha + \tilde{\lambda}'_\alpha \quad \forall \alpha \tag{7}$$

is sufficient to ensure the estimation $Y^{\mathrm{CK}}(\mathbf{x}_0)$ to be less than $Z^{\mathrm{CK}}(\mathbf{x}_0)$.

Because of both unbiasedness cokriging conditions $\sum_{\alpha=1}^{N}\lambda_\alpha = 1$ and $\sum_{\alpha=1}^{N}\lambda'_\alpha = 0$, we have:

$$\sum_{\alpha=1}^{N}\tilde{\lambda}'_\alpha \geq 1+\lambda_0 \quad \text{with } \lambda_0 \geq 0. \tag{8}$$

From the expression of $\tilde{\lambda}'_\alpha$ and inequality (8), we can use a similar framework to split and distribute a given quantity over the weights ν'_β, we have:

$$\sum_{\beta=1}^{M}\nu_\beta Z(\mathbf{x}_\beta) \leq \sum_{\beta=1}^{M}\left[\nu'_\beta - \tilde{\nu}'_\beta\right] Z(\mathbf{x}_\beta) \tag{9}$$

with:

$$\tilde{\nu}'_\beta = \frac{k(\mathbf{x}_\beta)(1+\lambda_0)}{Z(\mathbf{x}_\beta)\sum_{\alpha=1}^{N} l(\mathbf{x}_\alpha)/Y(\mathbf{x}_\alpha)},$$

and $\sum_{\beta=1}^{M}k(\mathbf{x}_\beta) = 1$. Again, if a uniform scheme distribution is used over the weights ν'_β, then $k(\mathbf{x}_\beta) = 1/M\ \forall \mathbf{x}_\beta$.

A sufficient condition to satisfy inequality (9) is:

$$\nu_\beta \leq \nu'_\beta - \tilde{\nu}'_\beta \quad \forall \mathbf{x}_\beta \tag{10}$$

And because of condition (10) and both unbiasedness conditions $\sum_{\beta=1}^{M}\nu_\beta = 0$ and $\sum_{\beta=1}^{M}\nu'_\beta = 1$:

$$\sum_{\beta=1}^{M}\tilde{\nu}'_\beta \leq 1-\nu_0 \quad \text{with } \nu_0 \geq 0 \tag{11}$$

Thus, for a given pair (λ_0, ν_0) satisfying conditions (8) and (11), the set of weights λ_α and ν_β compliant with inequality (6) have the following property:

$$\frac{1+\lambda_0}{\sum_\alpha l(\mathbf{x}_\alpha)/Y(\mathbf{x}_\alpha)} \leq \frac{1-\nu_0}{\sum_\beta k(\mathbf{x}_\beta)/Z(\mathbf{x}_\beta)} \tag{12}$$

Because the variables $Y(\mathbf{x})$ and $Z(\mathbf{x})$ are always positive, it is in most cases possible to find a subset of the observational data $\{Y(\mathbf{x}_\alpha), Z(\mathbf{x}_\beta)\}$ satisfying condition (12).

Regarding the other unbiasedness condition, Eqs. (13a,13b,13c,13d) appearing in the two cokriging estimations, i.e.:

$$\begin{cases} \sum_\alpha \lambda_\alpha f^i(\mathbf{x}_\alpha) + \sum_\beta \nu_\beta f^i(\mathbf{x}_\beta) = f^i(\mathbf{x}_0) & \text{(13a)} \\ \sum_\alpha \lambda'_\alpha f^i(\mathbf{x}_\alpha) + \sum_\beta \nu'_\beta f^i(\mathbf{x}_\beta) = f^i(\mathbf{x}_0) & \text{(13b)} \\ \sum_\beta \nu_\beta g^j(\mathbf{x}_\beta) = 0 & \text{(13c)} \\ \sum_\beta \nu'_\beta g^j(\mathbf{x}_\beta) = g^j(\mathbf{x}_0) & \text{(13d)} \end{cases}$$

In what follows, the notations $\overline{\lambda_\alpha}$, Eq. (14a), and $\overline{\nu_\beta}$, Eq. (14b), denote the penalized conditions (7) and (10):

$$\overline{\lambda_\alpha} = \lambda'_\alpha + \tilde{\lambda}'_\alpha \tag{14a}$$
$$\overline{\nu_\beta} = \nu'_\beta - \tilde{\nu}'_\beta \tag{14b}$$

2.2. The optimization problem

Here, we present the optimization problem $(\mathcal{P})$ derived from the formulation of cokriging with inequality constraints:

$$\begin{cases} \Omega = (\lambda, \nu) \\ \quad = \operatorname{argmin} \operatorname{Var}\left[Y^{\mathrm{CK}}(\mathbf{x}_0) - Y(\mathbf{x}_0)\right] | h_i(\Omega) = 0,\ g_j(\Omega) \leq 0 \\ \text{where } h_i(\Omega) = \begin{cases} \sum_\alpha \lambda_\alpha - 1 = 0 \\ \sum_\beta \nu_\beta = 0 \\ \sum_\alpha \lambda_\alpha f^i(\mathbf{x}_\alpha) + \sum_\beta \nu_\beta f^i(\mathbf{x}_\beta) = f^i_0 \\ \sum_\beta \nu_\beta g^j(\mathbf{x}_\beta) = 0 \end{cases} \\ \text{and } g_j(\Omega) = \begin{cases} \lambda_\alpha - \overline{\lambda_\alpha} \leq 0,\ \alpha = 1, \cdots, N \\ \nu_\beta - \overline{\nu_\beta} \leq 0,\ \beta = 1, \cdots, M \end{cases} \end{cases}$$

The Lagrangian of problem $(\mathcal{P})$ writes:

$$\begin{aligned} L(\Omega, \mu, \tau) &= \operatorname{Var}\left[Y^{\mathrm{CK}}(\mathbf{x}_0) - Y(\mathbf{x}_0)\right] + \sum_{i=1}^{l+p+2}\kappa_i h_i(\Omega) \\ &+ \sum_{j=1}^{2}\eta_j g_j(\Omega) \\ &= \operatorname{Var}\left[Y^{\mathrm{CK}}(\mathbf{x}_0) - Y(\mathbf{x}_0)\right] + 2\mu_0\left(\sum_\alpha \lambda_\alpha - 1\right) \\ &+ 2\upsilon_0\sum_\beta\left(\nu_\beta\right) + 2\mu_j\left(\sum_\beta \nu_\beta g^j(\mathbf{x}_\beta)\right) \\ &+ 2\upsilon_i\left(\sum_\alpha \lambda_\alpha f^i(\mathbf{x}_\alpha) + \sum_\beta \nu_\beta f^i(\mathbf{x}_\beta) - f^i_0\right) \\ &+ \sum_\alpha \tau_\alpha(\lambda_\alpha - \overline{\lambda_\alpha}) + \sum_\beta \omega_\beta(\nu_\beta - \overline{\nu_\beta}) \end{aligned}$$

If there is a local minimum Ω^* of $\mathrm{Var}\left[Y(\mathbf{x}_0) - Y^{\mathrm{CK}}(\mathbf{x}_0)\right]$, it exists $\boldsymbol{\kappa} = (\boldsymbol{\mu}, \boldsymbol{\upsilon})$, $\boldsymbol{\mu} \in \mathbb{R}^{l+1}$, $\boldsymbol{\upsilon} \in \mathbb{R}^{p+1}$, and $\boldsymbol{\eta} = (\boldsymbol{\tau}, \boldsymbol{\omega})$, $\boldsymbol{\tau} \in \mathbb{R}^N$, and $\boldsymbol{\omega} \in \mathbb{R}^M$ so that:

$$
\begin{cases}
\nabla_\Omega L(\Omega^*, \mu, \tau) = \begin{bmatrix} \dfrac{\partial L}{\partial \lambda_\alpha} \\ \dfrac{\partial L}{\partial \nu_\beta} \end{bmatrix} \\
= \begin{bmatrix} \displaystyle\sum_{\alpha'} \lambda_{\alpha'} C^1_{\alpha\alpha'} - C^1_{\alpha x} + \sum_\beta \nu_\beta C^{12}_{\alpha\beta} + \mu_0 + \upsilon_i f^i(\mathbf{x}_\alpha) + \tau_\alpha \\ \displaystyle\sum_{\beta'} \nu_{\beta'} C^2_{\beta\beta'} - C^{12}_{\beta x} + \sum_\alpha \lambda_\alpha C^{12}_{\alpha\beta} + \upsilon_0 + \mu_j g^j(\mathbf{x}_\beta) + \upsilon_i f^i(\mathbf{x}_\beta) + \omega_\beta \end{bmatrix} \\
= 0 \\
h_i(\Omega^*) = 0, \text{ i.e. } \begin{cases} \displaystyle\sum_\alpha \lambda_\alpha - 1 = 0 \\ \displaystyle\sum_\beta \nu_\beta = 0 \\ \displaystyle\sum_\alpha \lambda_\alpha f^i(\mathbf{x}_\alpha) + \sum_\beta \nu_\beta f^i(\mathbf{x}_\beta) = f^i_0 \\ \displaystyle\sum_\beta \nu_\beta g^j(\mathbf{x}_\beta) = 0 \end{cases} \\
\eta_j g_j(\Omega^*) = 0, \text{ i.e. } \begin{cases} \forall \alpha,\ \tau_\alpha(\lambda_\alpha - \overline{\lambda_\alpha}) = 0, \text{ i.e. } \displaystyle\sum_\alpha \tau_\alpha(\lambda_\alpha - \overline{\lambda_\alpha}) = 0 \\ \forall \beta,\ \omega_\beta(\nu_\beta - \overline{\nu_\beta}) = 0, \text{ i.e. } \displaystyle\sum_\beta \omega_\beta(\nu_\beta - \overline{\nu_\beta}) = 0 \end{cases} \\
\tau_\alpha \geq 0,\ \alpha = 1, \cdots, N \text{ and } \omega_\beta \geq 0,\ \beta = 1, \cdots, M
\end{cases}
$$

For sake of simplicity and generality, index 1 and 2 respectively denotes variable Y and Z, and the notation $C^1_{\alpha\alpha'}$ stands for $C_1(||\mathbf{x}_\alpha - \mathbf{x}_{\alpha'}||)$ in what follows. The same applies for the other covariances.

These are the so-called Karush-Kuhn-Tucker (KKT) conditions, see e.g. Rothenberg [32], that can be written in a more readable matrix form as follows:

$$
\begin{bmatrix} \mathbf{K} & \mathbf{A} & \mathbf{I} & 0 \\ \mathbf{A}' & 0 & 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{P} \\ \mathbf{M} \\ \tau_\alpha \\ \omega_\beta \end{bmatrix} = \begin{bmatrix} \mathbf{K_0} \\ \mathbf{F_0} \end{bmatrix} \tag{15a}
$$

where notations $\mathbf{K}$, $\mathbf{A}$, $\mathbf{P}$, $\mathbf{M}$, $\mathbf{K_0}$ and $\mathbf{F_0}$ are taken from Eq.(5) and with side constraints :

$$
\begin{aligned}
&\sum_\alpha \tau_\alpha(\lambda_\alpha - \overline{\lambda_\alpha}) = 0 \\
&\sum_\beta \omega_\beta(\nu_\beta - \overline{\nu_\beta}) = 0
\end{aligned} \tag{15b}
$$

$$
\begin{aligned}
&\lambda_\alpha - \overline{\lambda_\alpha} \leq 0 \\
&\nu_\beta - \overline{\nu_\beta} \leq 0
\end{aligned} \quad \text{and} \quad \begin{aligned} &\tau_\alpha \geq 0,\ \alpha = 1, \cdots, N \\ &\omega_\beta \geq 0,\ \beta = 1, \cdots, M \end{aligned} \tag{15c}
$$

This estimation of $Y(\mathbf{x}_0)$ obtained by solving this KKT-based penalization, Eqs. (15b) and (15c), of the cokriging system (15a) is now denoted $Y^{\mathrm{PCK}}(\mathbf{x}_0)$.

2.3. The complementary slackness conditions

Since the conditions (15c) requires all of the τ_α and ω_β be zero or positive while all the terms $(\lambda_\alpha - \overline{\lambda_\alpha})$ and $(\nu_\beta - \overline{\nu_\beta})$ be zero or negative, the condition (15b) requires :

$$
\begin{aligned}
\tau_\alpha = 0 \ \text{ or } \ \lambda_\alpha = \overline{\lambda_\alpha} \\
\text{and } \ \omega_\beta = 0 \ \text{ or } \ \nu_\beta = \overline{\nu_\beta}
\end{aligned}
$$

This is a particular case of complementary slackness condition, see e.g. Boyd and Vandenberghe [6], where the variables τ_α and ω_β are respectively the complementary pairs, also called Lagrangian multipliers or dual variables, corresponding to the constraints $\lambda_\alpha \leq \overline{\lambda_\alpha}$ and $\nu_\beta \leq \overline{\nu_\beta}$. Specifically, $\sum_\alpha \tau_\alpha(\lambda_\alpha - \overline{\lambda_\alpha}) = 0$ indicates that, at the optimal solution λ_α and without loss of generality, either τ_α is zero or $\lambda_\alpha = \overline{\lambda_\alpha}$, i.e. the inequality constraint is binding. The second constraint $\sum_\beta \omega_\beta(\nu_\beta - \overline{\nu_\beta}) = 0$ can be interpreted in a similar way.

In what follows, the positive kriging approach of Barnes and Johnson [2] is adapted to address our problem: for a given solution (not necessarily optimal), the cokriging weights satisfying condition (15c) are denoted as the **basic weights** λ_α and ν_β:

$$
\begin{aligned}
\mathcal{B}_\alpha &= \{\alpha, \lambda_\alpha < \overline{\lambda_\alpha},\ \tau_\alpha = 0\} \\
\mathcal{B}_\beta &= \{\beta, \nu_\beta < \overline{\nu_\beta},\ \omega_\beta = 0\}
\end{aligned} \tag{16}
$$

while the weights equal to their authorized upper-boundaries are identified as the **non-basic** weights:

$$\begin{aligned}\overline{\mathcal{B}_\alpha} &= \{\alpha, \lambda_\alpha = \overline{\lambda_\alpha},\ \tau_\alpha \geq 0\}\\ \overline{\mathcal{B}_\beta} &= \{\beta, \nu_\beta = \overline{\nu_\beta},\ \omega_\beta \geq 0\}\end{aligned} \tag{17}$$

Along this line, the set of τ_α and ω_β respectively complementary with the "basic" λ_α and ν_β will be called the "basic" τ_α and ω_β. The same idea applies for the definition of the non-basic τ_α and ω_β.

The penalized cokriging system for the estimation $Y^{\mathrm{PCK}}(\mathbf{x}_0)$ can be rearranged and partitioned by ordering rows and columns according to the basic and non-basic components of the solution:

$$\boldsymbol{\lambda} = \begin{bmatrix}\boldsymbol{\lambda}_\mathrm{b}\\ \boldsymbol{\lambda}_\mathrm{n}\end{bmatrix},\ \boldsymbol{\nu} = \begin{bmatrix}\boldsymbol{\nu}_\mathrm{b}\\ \boldsymbol{\nu}_\mathrm{n}\end{bmatrix},\ \boldsymbol{\tau} = \begin{bmatrix}\boldsymbol{\tau}_\mathrm{b}\\ \boldsymbol{\tau}_\mathrm{n}\end{bmatrix},\ \boldsymbol{\omega} = \begin{bmatrix}\boldsymbol{\omega}_\mathrm{b}\\ \boldsymbol{\omega}_\mathrm{n}\end{bmatrix}$$

where $\boldsymbol{\lambda}_\mathrm{b}$ denotes the set of basic weights $\lambda_\alpha,\ \alpha \in \mathcal{B}_\alpha$ and $\boldsymbol{\lambda}_\mathrm{n},\ \alpha \in \overline{\mathcal{B}_\alpha}$ the set of non-basic weights λ_α. The same type of notations are used for ν_β, τ_α and ω_β.

Reordering rows and columns with the basic and non-basic nomenclature, Eqs. (16) and (17), gives:

$$\begin{aligned}&\begin{bmatrix}\mathbf{K}_{11}^{\mathrm{bb}} & \mathbf{K}_{11}^{\mathrm{bn}} & \mathbf{K}_{12}^{\mathrm{bb}} & \mathbf{K}_{12}^{\mathrm{bn}}\\ \mathbf{K}_{11}^{\mathrm{nb}} & \mathbf{K}_{11}^{\mathrm{nn}} & \mathbf{K}_{12}^{\mathrm{nb}} & \mathbf{K}_{12}^{\mathrm{nn}}\\ \mathbf{K}_{21}^{\mathrm{bb}} & \mathbf{K}_{21}^{\mathrm{bn}} & \mathbf{K}_{22}^{\mathrm{bb}} & \mathbf{K}_{22}^{\mathrm{bn}}\\ \mathbf{K}_{21}^{\mathrm{nb}} & \mathbf{K}_{21}^{\mathrm{nn}} & \mathbf{K}_{22}^{\mathrm{nb}} & \mathbf{K}_{22}^{\mathrm{nn}}\end{bmatrix}\begin{bmatrix}\boldsymbol{\lambda}_\mathrm{b}\\ \boldsymbol{\lambda}_\mathrm{n}\\ \boldsymbol{\nu}_\mathrm{b}\\ \boldsymbol{\nu}_\mathrm{n}\end{bmatrix}\\ +&\begin{bmatrix}1 & 0 & 0 & \mathbf{f}_{1\mathrm{b}}^i\\ 1 & 0 & 0 & \mathbf{f}_{1\mathrm{n}}^i\\ 0 & \mathbf{g}_{2\mathrm{b}}^j & 1 & \mathbf{f}_{2\mathrm{b}}^i\\ 0 & \mathbf{g}_{2\mathrm{n}}^j & 1 & \mathbf{f}_{2\mathrm{n}}^i\end{bmatrix}\begin{bmatrix}\mu_0\\ \boldsymbol{\mu}_j\\ \upsilon_0\\ \boldsymbol{\upsilon}_i\end{bmatrix}\\ +&\begin{bmatrix}\mathbf{1}_{11}^{\mathrm{bb}} & 0_{11}^{\mathrm{bn}} & 0_{12}^{\mathrm{bb}} & 0_{12}^{\mathrm{bn}}\\ 0_{11}^{\mathrm{nb}} & \mathbf{1}_{11}^{\mathrm{nn}} & 0_{12}^{\mathrm{nb}} & 0_{12}^{\mathrm{nn}}\\ 0_{21}^{\mathrm{bb}} & 0_{21}^{\mathrm{bn}} & \mathbf{1}_{22}^{\mathrm{bb}} & 0_{22}^{\mathrm{bn}}\\ 0_{21}^{\mathrm{nb}} & 0_{21}^{\mathrm{nn}} & 0_{22}^{\mathrm{nb}} & \mathbf{1}_{22}^{\mathrm{nn}}\end{bmatrix}\begin{bmatrix}\boldsymbol{\tau}_\mathrm{b}\\ \boldsymbol{\tau}_\mathrm{n}\\ \boldsymbol{\omega}_\mathrm{b}\\ \boldsymbol{\omega}_\mathrm{n}\end{bmatrix} = \begin{bmatrix}\mathbf{K}_{10}^{\mathrm{b}}\\ \mathbf{K}_{10}^{\mathrm{n}}\\ \mathbf{K}_{120}^{\mathrm{b}}\\ \mathbf{K}_{120}^{\mathrm{n}}\end{bmatrix}\end{aligned} \tag{18a}$$

$$\begin{bmatrix}1 & 1 & 0 & 0\\ 0 & 0 & \mathbf{g}_{2\mathrm{b}}^j & \mathbf{g}_{2\mathrm{n}}^j\\ 0 & 0 & 1 & 1\\ \mathbf{f}_{1\mathrm{b}}^i & \mathbf{f}_{1\mathrm{n}}^i & \mathbf{f}_{2\mathrm{b}}^i & \mathbf{f}_{2\mathrm{n}}^i\end{bmatrix}\begin{bmatrix}\boldsymbol{\lambda}_\mathrm{b}\\ \boldsymbol{\lambda}_\mathrm{n}\\ \boldsymbol{\nu}_\mathrm{b}\\ \boldsymbol{\nu}_\mathrm{n}\end{bmatrix} = \begin{bmatrix}1\\ 0\\ 0\\ f_0^i\end{bmatrix} \tag{18b}$$

$$\begin{bmatrix}\boldsymbol{\tau}_\mathrm{b} & \boldsymbol{\tau}_\mathrm{n} & \boldsymbol{\omega}_\mathrm{b} & \boldsymbol{\omega}_\mathrm{n}\end{bmatrix}\begin{bmatrix}\boldsymbol{\lambda}_\mathrm{b} - \overline{\boldsymbol{\lambda}_\mathrm{b}}\\ \boldsymbol{\lambda}_\mathrm{n} - \overline{\boldsymbol{\lambda}_\mathrm{n}}\\ \boldsymbol{\nu}_\mathrm{b} - \overline{\boldsymbol{\nu}_\mathrm{b}}\\ \boldsymbol{\nu}_\mathrm{n} - \overline{\boldsymbol{\nu}_\mathrm{n}}\end{bmatrix} = 0 \tag{18c}$$

$$\begin{aligned}&\lambda_\alpha - \overline{\lambda_\alpha} \leq 0\\ &\nu_\beta - \overline{\nu_\beta} \leq 0\end{aligned} \quad \text{and} \quad \begin{aligned}&\tau_\alpha^* \geq 0,\ \alpha = 1, \cdots, N\\ &\omega_\beta^* \geq 0,\ \beta = 1, \cdots, M\end{aligned} \tag{18d}$$

$\mathbf{K}_{ij}^{\mathrm{bb}} = \{C_{ij}(\mathbf{x}_\alpha - \mathbf{x}'_\alpha)\},\ \alpha \in \mathcal{B}_\alpha,\ \alpha' \in \mathcal{B}_\alpha.$
$\mathbf{K}_{ij}^{\mathrm{bn}} = \{C_{ij}(\mathbf{x}_\alpha - \mathbf{x}'_\alpha)\},\ \alpha \in \mathcal{B}_\alpha,\ \alpha' \in \overline{\mathcal{B}_\alpha}.$
$\mathbf{K}_{ij}^{\mathrm{nb}} = \{C_{ij}(\mathbf{x}_\alpha - \mathbf{x}'_\alpha)\},\ \alpha \in \overline{\mathcal{B}_\alpha},\ \alpha' \in \mathcal{B}_\alpha.$
$\mathbf{K}_{ij}^{\mathrm{nn}} = \{C_{ij}(\mathbf{x}_\alpha - \mathbf{x}'_\alpha)\},\ \alpha \in \overline{\mathcal{B}_\alpha},\ \alpha' \in \overline{\mathcal{B}_\alpha}.$
$\mathbf{K}_{10}^{\mathrm{b}} = \{C_1(\mathbf{x}_\alpha - \mathbf{x}_0)\}, \alpha \in \mathcal{B}_\alpha.$
$\mathbf{K}_{10}^{\mathrm{n}} = \{C_1(\mathbf{x}_\alpha - \mathbf{x}_0)\}, \alpha \in \overline{\mathcal{B}_\alpha}.$
$\mathbf{K}_{120}^{\mathrm{b}} = \{C_{12}(\mathbf{x}_\beta - \mathbf{x}_0)\}, \beta \in \mathcal{B}_\beta.$
$\mathbf{K}_{120}^{\mathrm{n}} = \{C_{12}(\mathbf{x}_\beta - \mathbf{x}_0)\}, \beta \in \overline{\mathcal{B}_\beta}.$
$\mathbf{g}_{2\mathrm{b}}^j = \{g^j(\mathbf{x}_\beta)\}, \beta \in \mathcal{B}_\beta, \mathbf{g}_{2\mathrm{n}}^j = \{g^j(\mathbf{x}_\beta)\}, \beta \in \overline{\mathcal{B}_\beta}.$
$\mathbf{f}_{1\mathrm{b}}^i = \{f^i(\mathbf{x}_\alpha)\}, \alpha \in \mathcal{B}_\alpha, \mathbf{f}_{1\mathrm{n}}^i = \{f^i(\mathbf{x}_\alpha)\}, \alpha \in \overline{\mathcal{B}_\alpha}.$
$\mathbf{f}_{2\mathrm{b}}^i = \{f^i(\mathbf{x}_\beta)\}, \beta \in \mathcal{B}_\beta, \mathbf{f}_{2\mathrm{n}}^i = \{f^i(\mathbf{x}_\beta)\}, \beta \in \overline{\mathcal{B}_\beta}.$
$f_0^i = f^i(\mathbf{x}_0)$ is the ith covariate f^i at the target location $\mathbf{x}_0$.

Because the covariance function is assumed to be symmetric, see again Beauchamp et al. [3], $\mathbf{K}_{21}^{\mathrm{bb}}$,$\mathbf{K}_{21}^{\mathrm{bn}}$, $\mathbf{K}_{21}^{\mathrm{nb}}$ and $\mathbf{K}_{21}^{\mathrm{nn}}$ are respectively the transposes of the matrices $\mathbf{K}_{12}^{\mathrm{bb}}$, $\mathbf{K}_{12}^{\mathrm{bn}}$, $\mathbf{K}_{12}^{\mathrm{nb}}$ and $\mathbf{K}_{12}^{\mathrm{nn}}$.
$\mathbf{1}_{11}^{\mathrm{bb}}$ and $\mathbf{0}_{11}^{\mathrm{bb}}$ are the identity and zero matrices with same size than $\mathbf{K}_{11}^{\mathrm{bb}}$. The same applies for all the 16 possible combinations of covariance matrices between basic and non-basic weights of Y and Z.

Expanding the system (18a,18b,18c,18d) and because, by definition, the non-basic λ_α and ν_β respectively equals $\overline{\lambda_\alpha}$ and $\overline{\nu_\beta}$ and the basic τ_α and ω_β equals 0:

$$\begin{aligned}
&\mathbf{K}_{11}^{\mathrm{bb}}\lambda_{\mathrm{b}} && +\mathbf{K}_{12}^{\mathrm{bb}}\nu_{\mathrm{b}}+\mu_0+\upsilon_i\mathbf{f}_{1\mathrm{b}}^i \\
& && =\mathbf{K}_{10}^{\mathrm{b}}-\mathbf{K}_{11}^{\mathrm{bn}}\overline{\lambda_{\mathrm{b}}}-\mathbf{K}_{12}^{\mathrm{bn}}\overline{\nu_{\mathrm{b}}} && (19a)\\
&\mathbf{K}_{11}^{\mathrm{nb}}\lambda_{\mathrm{b}} && +\mathbf{K}_{12}^{\mathrm{nb}}\nu_{\mathrm{b}}+\mu_0+\upsilon_i\mathbf{f}_{1\mathrm{n}}^i+\mathbf{1}_{11}^{\mathrm{nn}}\boldsymbol{\tau}_{\mathrm{n}} \\
& && =\mathbf{K}_{10}^{\mathrm{n}}-\mathbf{K}_{11}^{\mathrm{nn}}\overline{\lambda_{\mathrm{b}}}-\mathbf{K}_{12}^{\mathrm{nn}}\overline{\nu_{\mathrm{b}}} && (19b)\\
&\mathbf{K}_{21}^{\mathrm{bb}}\lambda_{\mathrm{b}} && +\mathbf{K}_{22}^{\mathrm{bb}}\nu_{\mathrm{b}}+\upsilon_0+\boldsymbol{\mu}_j\mathbf{g}_{2\mathrm{b}}^j+\boldsymbol{\upsilon}_i\mathbf{f}_{2\mathrm{b}}^i \\
& && =\mathbf{K}_{20}^{\mathrm{b}}-\mathbf{K}_{21}^{\mathrm{bn}}\overline{\lambda_{\mathrm{b}}}-\mathbf{K}_{22}^{\mathrm{bn}}\overline{\nu_{\mathrm{b}}} && (19c)\\
&\mathbf{K}_{21}^{\mathrm{nb}}\lambda_{\mathrm{b}} && +\mathbf{K}_{22}^{\mathrm{nb}}\nu_{\mathrm{b}}+\upsilon_0+\boldsymbol{\mu}_j\mathbf{g}_{2\mathrm{n}}^j+\boldsymbol{\upsilon}_i\mathbf{f}_{2\mathrm{n}}^i+\mathbf{1}_{22}^{\mathrm{nn}}\boldsymbol{\omega}_{\mathrm{n}} \\
& && =\mathbf{K}_{20}^{\mathrm{n}}-\mathbf{K}_{21}^{\mathrm{nn}}\overline{\lambda_{\mathrm{b}}}-\mathbf{K}_{22}^{\mathrm{nn}}\overline{\nu_{\mathrm{b}}} && (19d)\\
&\mathbf{1}^{\mathrm{b}}\lambda_{\mathrm{b}} && =1-\mathbf{1}^{\mathrm{n}}\overline{\lambda_{\mathrm{n}}} && (19e)\\
&\mathbf{g}_{2\mathrm{b}}^j\lambda_{\mathrm{b}} && =-\mathbf{g}_{2\mathrm{n}}^j\overline{\nu_{\mathrm{n}}} && (19f)\\
&\mathbf{1}^{\mathrm{b}}\nu_{\mathrm{b}} && =-\mathbf{1}^{\mathrm{n}}\overline{\nu_{\mathrm{n}}} && (19g)\\
&\mathbf{f}_{1\mathrm{b}}^i\lambda_{\mathrm{b}}+\mathbf{f}_{2\mathrm{b}}^j\nu_{\mathrm{b}} && =f_0^i-\mathbf{f}_{1\mathrm{n}}^i\lambda_{\mathrm{n}}-\mathbf{f}_{2\mathrm{n}}^j\overline{\nu_{\mathrm{n}}} && (19h)
\end{aligned}$$

$$\begin{aligned}
&\lambda_\alpha-\overline{\lambda_\alpha}\le 0 \\
&\nu_\beta-\overline{\nu_\beta}\le 0
\end{aligned}
\quad\text{and}\quad
\begin{aligned}
&\tau_\alpha^*\ge 0,\ \alpha=1,\cdots,N \\
&\omega_\beta^*\ge 0,\ \beta=1,\cdots,M
\end{aligned}
\qquad (19i)$$

The condition (18c) is implicitly satisfied at optimality by the definitions of the basic and non-basic λ, ν, τ and ω.

Rewriting conditions (19a), (19c), (19e), (19f), (19g), (19h) in a convenient matrix form (non-basic conditions (19b) and (19d) are not embedded in the linear system) leads to:

$$\begin{bmatrix}
\mathbf{K}_{11}^{\mathrm{bb}} & \mathbf{K}_{12}^{\mathrm{bb}} & 1 & 0 & 0 & \mathbf{f}_{1\mathrm{b}}^i \\
\mathbf{K}_{21}^{\mathrm{bb}} & \mathbf{K}_{22}^{\mathrm{bb}} & 0 & \mathbf{g}_{2\mathrm{b}}^j & 1 & \mathbf{f}_{2\mathrm{b}}^i \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & \mathbf{g}_{2\mathrm{b}}^j & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
\mathbf{f}_{1\mathrm{b}}^i & \mathbf{f}_{2\mathrm{b}}^i & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\lambda_{\mathrm{b}} \\ \nu_{\mathrm{b}} \\ \mu_0 \\ \boldsymbol{\mu}_j \\ \upsilon_0 \\ \boldsymbol{\upsilon}_i
\end{bmatrix}$$

$$=\begin{bmatrix}
\mathbf{K}_{10}^{\mathrm{b}}-\mathbf{K}_{11}^{\mathrm{bn}}\overline{\lambda_{\mathrm{b}}}-\mathbf{K}_{12}^{\mathrm{bn}}\overline{\nu_{\mathrm{b}}} \\
\mathbf{K}_{20}^{\mathrm{n}}-\mathbf{K}_{21}^{\mathrm{bn}}\overline{\lambda_{\mathrm{b}}}-\mathbf{K}_{22}^{\mathrm{bn}}\overline{\nu_{\mathrm{b}}} \\
1-\mathbf{1}^{\mathrm{n}}\overline{\lambda_{\mathrm{n}}} \\
-\mathbf{g}_{2\mathrm{n}}^j\overline{\nu_{\mathrm{n}}} \\
-\mathbf{1}^{\mathrm{n}}\overline{\nu_{\mathrm{n}}} \\
f_0^i-\mathbf{f}_{1\mathrm{n}}^i\overline{\lambda_{\mathrm{n}}}-\mathbf{f}_{2\mathrm{n}}^i\overline{\nu_{\mathrm{n}}}
\end{bmatrix}$$

The final estimation of $Y(\mathbf{x}_0)$ will be:

$$\begin{aligned}
Y^{\mathrm{PCK}}(\mathbf{x}_0) \quad &= \sum_{\alpha\in\mathcal{B}_\alpha}\lambda_\alpha Y(\mathbf{x}_\alpha)+\sum_{\alpha\in\overline{\mathcal{B}_\alpha}}\overline{\lambda_\alpha}Y(\mathbf{x}_\alpha) \\
&+\sum_{\beta\in\mathcal{B}_\beta}\nu_\beta Z(\mathbf{x}_\beta)+\sum_{\beta\in\overline{\mathcal{B}_\beta}}\overline{\nu_\beta}Z(\mathbf{x}_\beta) \quad (20)
\end{aligned}$$

with $\mathcal{B}_\alpha$ and $\mathcal{B}_\beta$ the set of index for the basic weights of $Y(\mathbf{x}_\alpha)$ and $Z(\mathbf{x}_\beta)$, and $\overline{\mathcal{B}_\alpha}$ and $\overline{\mathcal{B}_\beta}$ their complementary, see Eq.(16) and (17).

The related cokriging variance of $Y^{\mathrm{PCK}}(\mathbf{x}_0)$, Eq. (20), is (see the detailed calculations in AppendixA):

$$\begin{aligned}
&\mathrm{Var}\left[Y(\mathbf{x}_0)-Y^{\mathrm{PCK}}(\mathbf{x}_0)\right]=\sigma^2_{\mathrm{PCK}}(\mathbf{x}_0) \\
&=C^1(\mathbf{0})-\sum_{\alpha\in\mathcal{B}_\alpha}\lambda_\alpha C^1_{\alpha 0}-\sum_{\alpha\in\overline{\mathcal{B}_\alpha}}\lambda_\alpha\left[C^1_{\alpha 0}-\tau_\alpha\right] \\
&-\sum_{\beta\in\mathcal{B}_\beta}\nu_\beta C^{12}_{\beta 0}-\sum_{\beta\in\overline{\mathcal{B}_\beta}}\nu_\beta\left[C^{12}_{\beta 0}-\omega_\beta\right]-\mu_0-\sum_i\upsilon_i f^i(\mathbf{x}_0)
\end{aligned}$$

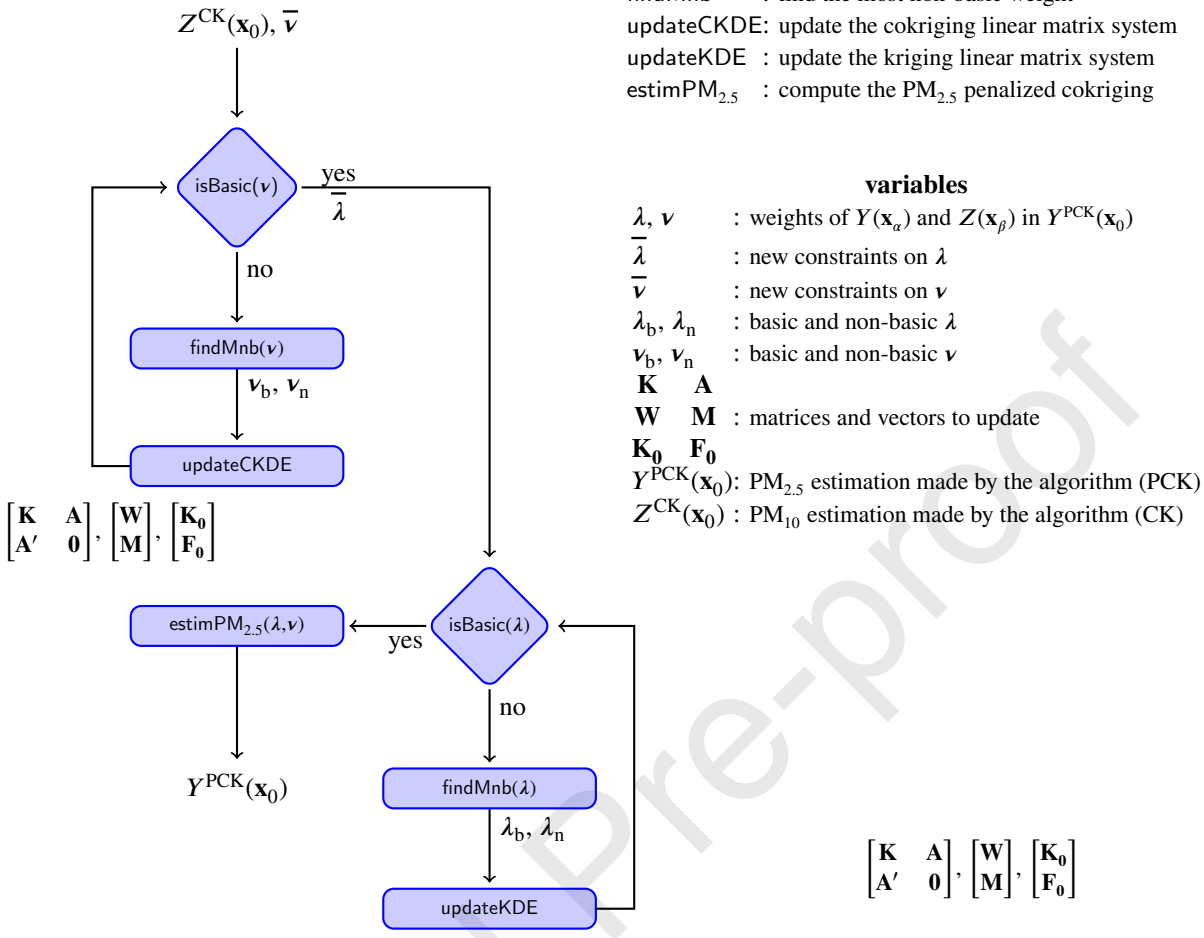
where $\{\tau_\alpha\}$ and $\{\omega_\beta\}$ are directly given by Eq.(19b) and (19d).

2.4. Strategy

The basic and non-basic λ_α satisfy the condition (7), which requires to know the basic and non-basic ν_β first, satisfying the condition (10). The first step thus consists in solving the cokriging system looking at each step of the algorithm which ν_β is the most non-basic, see section 2.6 and following Eq. (21), until there is no non-basic ν_β anymore.

$$\begin{bmatrix}
\mathbf{K}_{11}^{\mathrm{bb}} & \mathbf{K}_{12}^{\mathrm{bb}} & 1 & 0 & 0 & \mathbf{f}_{1\mathrm{b}}^i \\
\mathbf{K}_{21}^{\mathrm{bb}} & \mathbf{K}_{22}^{\mathrm{bb}} & 0 & \mathbf{g}_{2\mathrm{b}}^j & 1 & \mathbf{f}_{2\mathrm{b}}^i \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & \mathbf{g}_{2\mathrm{b}}^j & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
\mathbf{f}_{1\mathrm{b}}^i & \mathbf{f}_{2\mathrm{b}}^i & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\lambda_{\mathrm{b}} \\ \nu_{\mathrm{b}} \\ \mu_0 \\ \boldsymbol{\mu}_j \\ \upsilon_0 \\ \boldsymbol{\upsilon}_i
\end{bmatrix}
=\begin{bmatrix}
\mathbf{K}_{10}^{\mathbf{b}}-\mathbf{K}_{12}^{\mathrm{bn}}\overline{\nu_{\mathrm{b}}} \\
\mathbf{K}_{20}^{\mathbf{n}}-\mathbf{K}_{22}^{\mathrm{bn}}\overline{\nu_{\mathrm{b}}} \\
1 \\
-\mathbf{g}_{2\mathrm{n}}^j\overline{\nu_{\mathrm{n}}} \\
-\mathbf{1}^{\mathrm{n}}\overline{\nu_{\mathrm{n}}} \\
f_0^i-\mathbf{f}_{2\mathrm{n}}^i\overline{\nu_{\mathrm{n}}}
\end{bmatrix} \quad (21)$$

Then, the ν_β are considered as known and fixed. The cok-



subprocedures

isBasic	: test if all the weights are basic
findMnb	: find the most non-basic weight
updateCKDE	: update the cokriging linear matrix system
updateKDE	: update the kriging linear matrix system
estimPM$_{2.5}$	: compute the PM$_{2.5}$ penalized cokriging

variables

λ, ν	: weights of $Y(\mathbf{x}_\alpha)$ and $Z(\mathbf{x}_\beta)$ in $Y^{PCK}(\mathbf{x}_0)$
$\overline{\lambda}$	: new constraints on λ
$\overline{\nu}$	: new constraints on ν
λ_b, λ_n	: basic and non-basic λ
ν_b, ν_n	: basic and non-basic ν
$\mathbf{K}$ $\mathbf{A}$ $\mathbf{W}$ $\mathbf{M}$ $\mathbf{K_0}$ $\mathbf{F_0}$	: matrices and vectors to update
$Y^{PCK}(\mathbf{x}_0)$	: PM$_{2.5}$ estimation made by the algorithm (PCK)
$Z^{CK}(\mathbf{x}_0)$	: PM$_{10}$ estimation made by the algorithm (CK)

Figure 1: algorigram of the penalized cokriging (PCK) algorithm

riging system is after that modified to satisfy the condition (7). The terms related to the second variable Z are moved to the right-hand side of the linear system, so that the unbiasedness conditions and optimality are still satisfied:

$$\begin{bmatrix}\mathbf{K}^{bb}_{11} & 1 & \mathbf{f}^{i}_{1b}\\ 1 & 0 & 0\\ \mathbf{f}^{i}_{1b} & 0 & 0\end{bmatrix}\begin{bmatrix}\lambda_b\\ \mu_0\\ \upsilon_i\end{bmatrix}=\begin{bmatrix}\mathbf{K}^{b}_{10}-\mathbf{K}^{bn}_{11}\overline{\lambda_n}-\mathbf{K}^{bn}_{12}\overline{\nu_n}-\mathbf{K}^{bb}_{12}\overline{\nu_b}\\ 1-\mathbf{1}^{n}\overline{\lambda_n}\\ f^{i}_{0}-\mathbf{f}^{i}_{1n}\overline{\lambda_n}-\mathbf{f}^{i}_{2n}\overline{\nu_n}-\mathbf{f}^{i}_{2b}\overline{\nu_b}\end{bmatrix} \quad (22)$$

The algorithm may still reduce the set of basic weights too far and the estimation will just be a heuristic produced by the additional constraints made on the cokriging system. Because two unbiasedness conditions are still appearing in the system (22), two basic weights are at least required.

2.5. Computational details

Given that $\mathbf{A}$ is symmetric, $\mathbf{A}^{-1}$ exists and the inverse of the matrix $\mathbf{A}$ extended by an additional row and column is:

$$\begin{bmatrix}\mathbf{A} & \mathbf{a}\\ \mathbf{a}' & \alpha\end{bmatrix}^{-1}=\begin{bmatrix}\mathbf{C} & \mathbf{c}\\ \mathbf{c}' & -\beta\end{bmatrix}$$

Thus, $\mathbf{A}^{-1}$ can be efficiently computed: $\mathbf{A}^{-1}=\mathbf{C}+(1/\beta)\mathbf{c}\mathbf{c}'$.

When classifying a weight λ_α or ν_β as non-basic, the row and the column to remove are generally not the last ones. The tool to compute the inverse of a matrix $\mathbf{A}$ after a permutation of the row and column j to the far right and bottom of matrix $\mathbf{A}$ is also given in [2]:

$$\begin{bmatrix}a_{11} & \dots & a_{1j} & \dots & a_{1n}\\ \vdots & & \vdots & & \\ a_{j1} & \dots & a_{jj} & \dots & a_{jn}\\ \vdots & & \vdots & & \\ a_{n1} & \dots & a_{nj} & \dots & a_{nn}\end{bmatrix}^{-1}=\begin{bmatrix}c_{11} & \dots & c_{1j} & \dots & c_{1n}\\ \vdots & & \vdots & & \\ c_{j1} & \dots & c_{jj} & \dots & c_{jn}\\ \vdots & & \vdots & & \\ c_{n1} & \dots & c_{nj} & \dots & c_{nn}\end{bmatrix}$$

$$\Rightarrow\left[\begin{array}{ccc|c}a_{11} & \dots & a_{1n} & a_{1j}\\ \vdots & & \vdots & \vdots\\ a_{n1} & & a_{nn} & a_{nj}\\ \hline a_{j1} & & a_{jn} & a_{jj}\end{array}\right]^{-1}=\left[\begin{array}{ccc|c}c_{11} & \dots & c_{1n} & c_{1j}\\ \vdots & & \vdots & \vdots\\ c_{n1} & & c_{nn} & c_{nj}\\ \hline c_{j1} & & c_{jn} & c_{jj}\end{array}\right] \quad (23)$$

Instead of solving the new linear system at each iteration of the algorithm, it is possible with the previously defined

tools to quickly update the inverse of the matrix and simply multiply it with the update of the second member. An even faster solution is to use some simple properties of linear algebra. Let us note $\begin{bmatrix}\mathbf{y} & z\end{bmatrix}'$ the solution of the linear system at step K of the algorithm. Thus,

$$\begin{bmatrix}\mathbf{A} & \mathbf{a}\\ \mathbf{a}' & \alpha\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{K}\\ k\end{bmatrix}=\begin{bmatrix}\mathbf{C} & \mathbf{c}\\ \mathbf{c}' & -\beta\end{bmatrix}\begin{bmatrix}\mathbf{K}\\ k\end{bmatrix}=\begin{bmatrix}\mathbf{y}\\ z\end{bmatrix}$$

At step $K+1$, a new observation is defined as non-basic. The new basic weights are obtained by multiplying the matrix $\mathbf{A}^{-1}$ with the column vector $\mathbf{K}-\mathbf{a}\overline{z}$, where $\overline{z}$ is the non-basic weights vector related to z, i.e.:

$$\begin{aligned}\mathbf{A}^{-1}(\mathbf{K}-\mathbf{a}\overline{z}) &= \mathbf{C}+(\frac{1}{\beta})\mathbf{c}\mathbf{c}'(\mathbf{K}-a\overline{z})\\ &= \mathbf{C}\mathbf{K}+\frac{1}{\beta}\mathbf{c}\mathbf{c}'\mathbf{K}-\mathbf{C}\mathbf{a}\overline{z}-\frac{1}{\beta}\mathbf{c}\mathbf{c}'\mathbf{a}\overline{z}\\ &= \mathbf{y}+\frac{z}{\beta}\mathbf{c}-\mathbf{C}\mathbf{a}\overline{z}-\frac{1}{\beta}\mathbf{c}\mathbf{c}'\mathbf{a}\overline{z}\end{aligned}$$

By noticing that:

$$\begin{bmatrix}\mathbf{C} & \mathbf{c}\\ \mathbf{c}' & -\beta\end{bmatrix}\begin{bmatrix}\mathbf{A} & \mathbf{a}\\ \mathbf{a}' & \alpha\end{bmatrix}^{-1}=\mathbf{I}$$

implies that:

$$\begin{bmatrix}\mathbf{C} & \mathbf{c}\end{bmatrix}\begin{bmatrix}\mathbf{a}\\ \alpha\end{bmatrix} =0, \text{ i.e. } \mathbf{C}\mathbf{a}\overline{z}=-\alpha\mathbf{c}\overline{z}$$

$$\begin{bmatrix}\mathbf{c}' & -\beta\end{bmatrix}\begin{bmatrix}\mathbf{a}\\ \alpha\end{bmatrix} =1, \text{ i.e. } \mathbf{d}'\mathbf{a}\overline{z}=(1+\beta\alpha)\overline{z}$$

Finally,

$$\begin{aligned}\mathbf{A}^{-1}(\mathbf{K}-\mathbf{a}\overline{z}) &= (\mathbf{C}+(\frac{1}{\beta})\mathbf{c}\mathbf{c}')(\mathbf{K}-a\overline{z})\\ &= \mathbf{y}+\mathbf{c}\left(\frac{z-\overline{z}(1+\alpha(\beta+1))}{\beta}\right)\end{aligned}$$

The computational cost is thus considerably reduced. Instead of multiplying a squared matrix of rank $N(\mathcal{B})=\#\{\mathcal{B}_\alpha\cup\mathcal{B}_\beta\}$, i.e. the number of basic weights, with a column vector of size $N(\mathcal{B})$, that is $N(\mathcal{B})^2$ multiplications and $[N(\mathcal{B})-1]^2$ additions, the numbers of operations required to update the basic weights is now only $N(\mathcal{B})$+3 additions and $N(\mathcal{B})$+3 multiplications.

2.6. Remarks

A) The computational time of the method presented here may remain quite restricting when using a unique neighborhood because cokriging matrices have a size much bigger than kriging matrices. Thus, despite the previous tools introduced in Sect. 2.5 to avoid the solving of the linear system at each step of the algorithm, the update of the second member, the matrix cokriging and the weights may still be too costly for an operational algorithm. A solution to speed up the estimation can be to consider a moving neighborhood and is fully detailed in Appendix B.

B) At each step of the algorithm, if there are still some weights that do not satisfy the constraints (7), the "most" non-basic weights λ_α have to be removed from the set of basic weights. The most non-basic index α can be defined by :
Eq. (24), as the weight that satisfies the least the constraints (i.e that maximizes their differences):

$$\underset{\alpha}{\text{argmax}}\{\lambda_\alpha-\overline{\lambda_\alpha}>0\} \tag{24}$$

or as Eq. (25), as the weight non satisfying the constraints (7) and maximizing the distance with the target location $\mathbf{x}_0$ where an estimation has to be made:

$$\underset{\alpha}{\text{argmax}}\{||\mathbf{x}_\alpha-\mathbf{x}_0||,\ \lambda_\alpha-\overline{\lambda_\alpha}>0\} \tag{25}$$

This second option allows the weights related to the observations in a close neighborhood of $\mathbf{x}_0$ to be truly estimated and not set to their non-basic related values when the algorithm successively reduces the set of basic weights. The same applies to find the most non-basic weights ν_β.

3. Results

In this Section, we proposed an application based on the dataset used in Beauchamp et al. [3]. The specific additive modelling introduced in the above-mentioned work is used to compare both mappings and performance of the penalized cokriging (PCK) algorithm with the usual way of solving the cokriging system (CK). The background PM_{10} and background $PM_{2.5}$ observational dataset is then identical to Beauchamp et al. [3], but extended from France to Europe, see Table 1, because the discontinuities discussed in Appendix B are more frequently seen when the domain is larger.

	Rural	Suburban	Urban
PM_{10}	190	183	445
$PM_{2.5}$	63	51	188

Table 1: Number of background monitoring sites (PM_{10} and $PM_{2.5}$)

The CHIMERE PM_{10} and $PM_{2.5}$ simulations are used as covariates for $m_Z(\mathbf{x})$ and $m_Y(\mathbf{x})$. The model covers the AWM European domain of simulation ($[-15°W, 35°W; 35°N, 70°N]$), with a coarse resolution of 0.5°. The maps are interpolated by cokriging on a regular grid with the same resolution used by CHIMERE for solving the chemical and physical processes.

The time period covers the first quarter of 2015 in which the PM concentrations were particularly high with a long episode of pollution occuring in March 2015.

Regarding the practical implementation, an original C++ program has been developed to run the penalized cokriging version. The armadillo C++ library [12] is involved for the numerical analysis related to kriging. An interface with R software [29] is used inside the C++ program towards the R library RCppArmadillo [17] so that all the descriptive geostatistical part, e.g. the variogram computation and fitting, are done by RGeostats through this interface. The full code is available via Zenodo (10.5281/zenodo.7756425). It is ready-to-use after installation of the appropriate libraries and specifications of the input data. The appropriate shape for the inputs is given as comment lines along the program.

3.1. Mapping

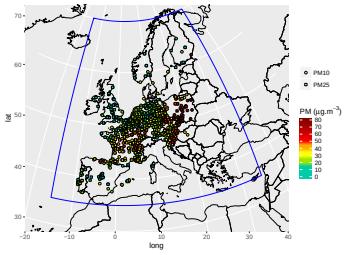
On the 10^{th} of March, Fig. 2 (a) to Fig. 2 (j) respectively show the daily observations, the corresponding scatterplot between the collocated $PM_{2.5}$ and PM_{10} sites, the CHIMERE simulations for the two pollutants and the kriging maps with their related standard deviations for PM_{10} and $PM_{2.5}$ with the classic algorithm (CK) and its penalized version (PCK). A zoomed-in window is given on Hungary, Serbia, Ukraine and Slovakia to focus on this area where $Y^{\mathrm{CK}}(\mathbf{x}_0) > Z^{\mathrm{CK}}(\mathbf{x}_0)$. The estimations $Y^{\mathrm{PCK}}(\mathbf{x}_0)$ are successfully less than $Z^{\mathrm{CK}}(\mathbf{x}_0)$. Looking at the daily-averaged observations in this area, that are only two in Hungary, the algorithm (PCK) seems to extend the representativeness areas of these two monitoring sites over the whole area where the inconsistency appears. PM_{10} observations are in the range 30-40 $\mu g.m^{-3}$ while $PM_{2.5}$ cokriging estimates values are greater than 40 $\mu g.m^{-3}$ over the area. Adding the penalization enables to decrease $PM_{2.5}$ estimations down to 30-40 $\mu g.m^{-3}$ which is not unrealistic when regarding how the values are spatially distributed. Though, an additional assessment of the true representativeness areas of these sites would be necessary to conclude. The same type of results is shown on the 16^{th} of March in Fig. 3 (a) to Fig. 3 (j), that is a typical example of far-off extrapolations in Northern Europe where physical inconsistencies can also occur. Once again, the algorithm (PCK) is successful and enables to decrease the prior values produced by the algorithm (CK) from 30 $\mu g.m^{-3}$ to less than 10 $\mu g.m^{-3}$ in its (PCK) version. It is to note that standard deviation of (PCK) errors were expected to be greater than those produced without the penalization but in most cases, as supported by the two examples given, they are in the same range of values for both algorithms.

More generally, the differences are often not significant from a mapping perspective. It was expected since the usual cokriging algorithm does not generate so many $PM_{2.5}$ estimations that are greather than their corresponding PM_{10} estimations. On some days however, the differences between the cokriging and its penalized version are significant, especially because the algorithm (PCK) has the direct consequence to fix the (non-basic) weights of a some observations in the estimation process. As a consequence, even if all the available observations are first used as input data, the progressive decrease of the set of basic weights leads to a map that seems to be built with a moving neighborhood. This one has very specific features strongly depending on the strategy used to define the most non-basic weight, see Sect. 2.6:

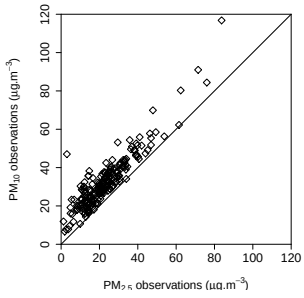
1) if it is defined as the farthest non-basic weight from the target point $\mathbf{x}_0$, see Eq.(25), then all the observations with basic weights are included in some distance-based neighborhood. If this distance is large enough, which is generally the case, the approach proposed in Appendix B makes vanish the discontinuities because the farthest observations, although with fixed (non-basic) weights, are noisy when estimating the PM_{10} concentration $Z(\mathbf{x_0})$.
2) if it is defined as the most non-basic weight, i.e. the weight with the largest positive deviation to its related non-basic version, see Eq.(24), then the set of basic weights fail to comply with some distance-based neighborhood. As a consequence, the estimation process still may generate strong discontinuities. It would clearly be the best option to keep the largest set of basic weights, but for mapping concerns, the way of dealing with these discontinuities appears problematic and Appendix B will not really help.

3.2. Cross-validation

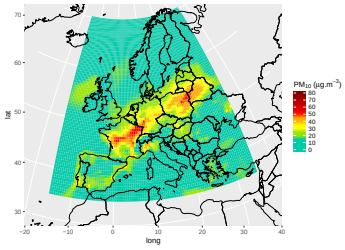
Figure 5 shows the results of a (leave-one-out) cross-validation procedure carried out every day of the first quarter of 2015. Figure 5a presents the scatterplot of the $PM_{2.5}$ observations (X-axis) and the $PM_{2.5}$ estimations (Y-axis) obtained by the cokriging algorithm (CK) of Beauchamp et al. [3] and its penalized version (PCK) detailed in Sect. 2. In the latter, the definition (25) of what should be the most non-basic weight is used. Over the period, only the observations $Y(\mathbf{x}_\alpha)$ with inconsistent cross-validation values $Y^{\mathrm{CK}}(\mathbf{x}_\alpha) > Z^{\mathrm{CK}}(\mathbf{x}_\alpha)$ are displayed and used to compute the correlation and RMSE. Let precise that observations sites with inconsistent cross-validation values have a significant number of occurences, see Figure 4 focusing on the French domain and same evaluation period, with related inconsistencies on classic cokriging mapping (without additive model) also noticeable, see the background mapping on the above-mentioned Figure.



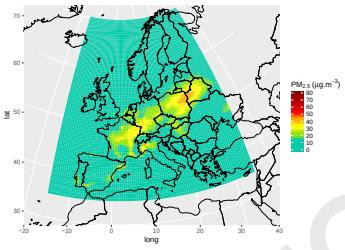
(a) Observations



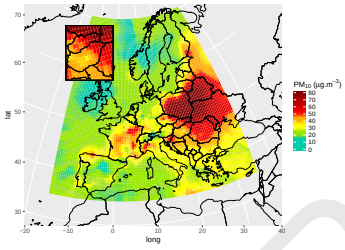
(b) Scatterplot $PM_{2.5}$ and PM_{10}



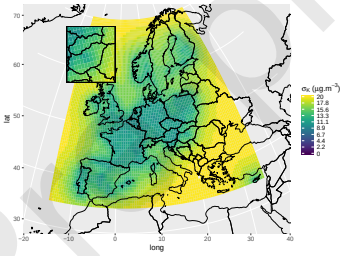
(c) CHIMERE (PM_{10})



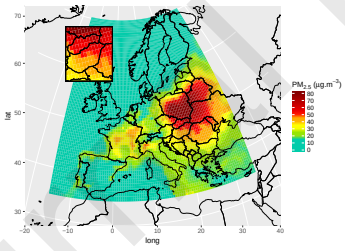
(d) CHIMERE ($PM_{2.5}$)



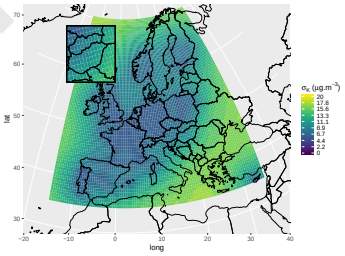
(e) PM_{10} Cokriging (CK)



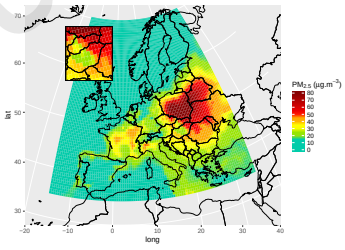
(f) PM_{10} standard deviations (CK)



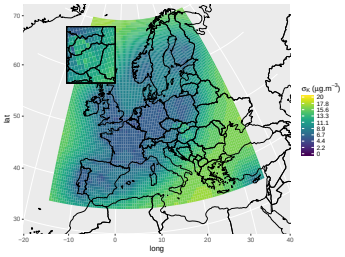
(g) $PM_{2.5}$ cokriging (CK)



(h) $PM_{2.5}$ standard deviations (CK)

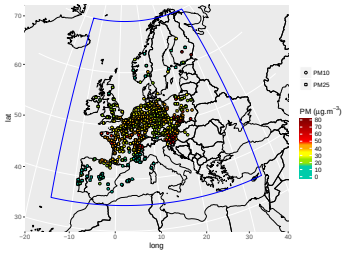


(i) $PM_{2.5}$ penalized cokriging (PCK)

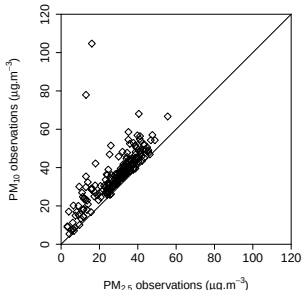


(j) $PM_{2.5}$ standard deviations (PCK)

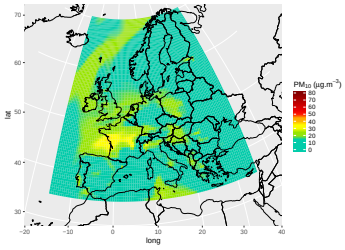
Figure 2: From top to bottom, PM_{10} and $PM_{2.5}$ available observations and corresponding scatterplots, PM_{10} and $PM_{2.5}$ CHIMERE outputs, PM_{10} cokriging and standard deviation as implemented in [3], the same for $PM_{2.5}$ and last, the proposed $PM_{2.5}$ penalized cokriging and standard deviation as proposed in this paper (2015, March 10). A focus is provided on a small box region over Hungary, Serbia, Ukraine and Slovakia with inconsistent cokriging estimations



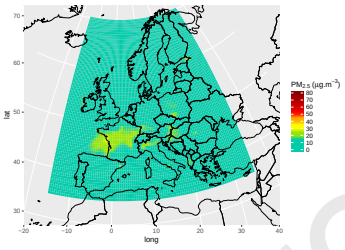
(a) Observations



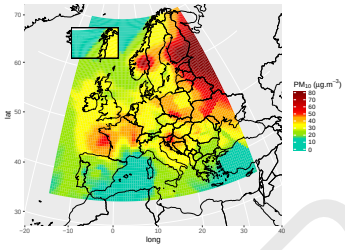
(b) Scatterplot $PM_{2.5}$ and PM_{10}



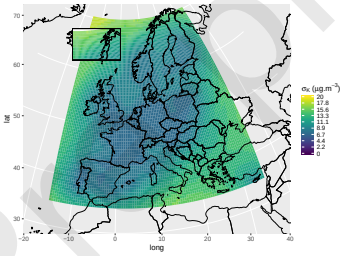
(c) CHIMERE (PM_{10})



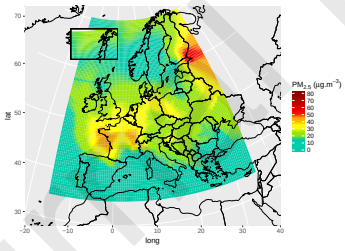
(d) CHIMERE ($PM_{2.5}$)



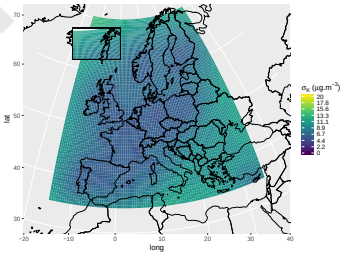
(e) PM_{10} Cokriging (CK)



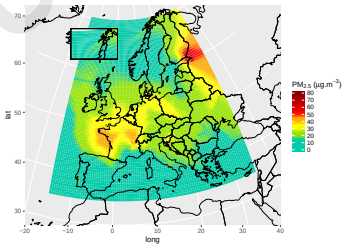
(f) PM_{10} standard deviations (CK)



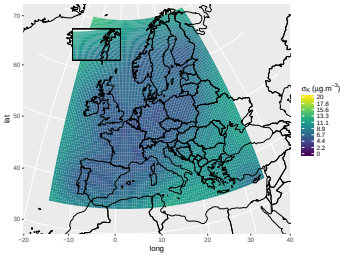
(g) $PM_{2.5}$ cokriging (CK)



(h) $PM_{2.5}$ standard deviations (CK)

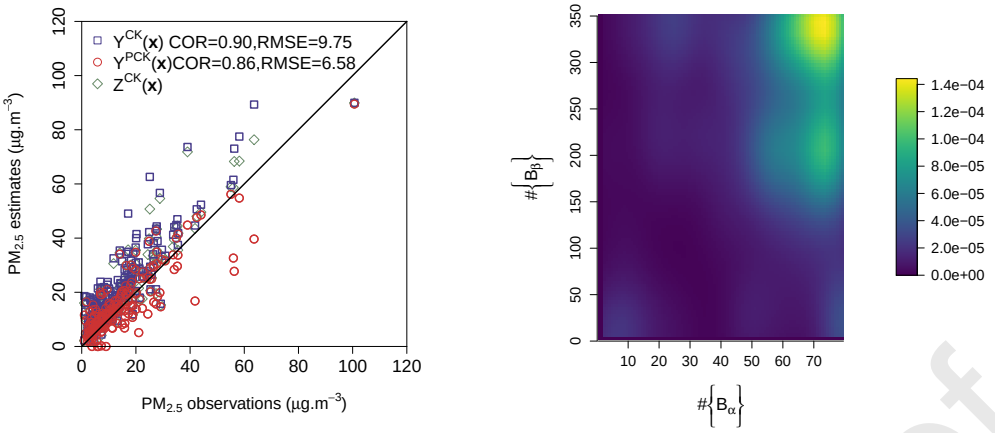


(i) $PM_{2.5}$ penalized cokriging (PCK)



(j) $PM_{2.5}$ standard deviations (PCK)

Figure 3: From top to bottom, PM_{10} and $PM_{2.5}$ available observations and corresponding scatterplots, PM_{10} and $PM_{2.5}$ CHIMERE outputs, PM_{10} cokriging and standard deviation as implemented in [3], the same for $PM_{2.5}$ and last, the proposed $PM_{2.5}$ penalized cokriging and standard deviation as proposed in this paper (2015, March 16). A focus is provided on a small box region o the western coastal area of Norway with inconsistent cokriging estimations



(a) Scatterplot $PM_{2.5}$ observations (X-axis) and $PM_{2.5}$ estimations (Y-axis)

(b) Bivariate distribution of $\#\{\mathcal{B}_\alpha\}$ and $\#\{\mathcal{B}_\beta\}$

Figure 5: Cross-validation procedure over the first quarter of 2015

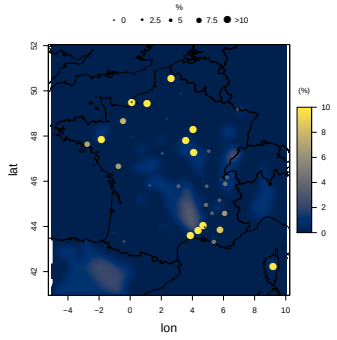


Figure 4: Frequencies of $PM_{2.5}/PM_{10}$ ratio greater than 1 for both observation sites (cross-validation) and cokriging estimation maps

In the case of consistent cokriging-based estimations, the scores of the two methods are similar.

$$1)\ \text{correlation} \quad : \quad \frac{\mathrm{Cov}\left[(Z(\mathbf{x}_\alpha), Z^{\mathrm{CK}}(\mathbf{x}_\alpha)\right]}{\sigma\left[Z(\mathbf{x}_\alpha)\right]\sigma\left[Z^{\mathrm{CK}}(\mathbf{x}_\alpha)\right]}$$

$$2)\ \text{(RMSE)} \quad : \quad \sqrt{\frac{1}{n}\sum_{\alpha=1}^{n}\left[Z(\mathbf{x}_\alpha) - Z^{\mathrm{CK}}(\mathbf{x}_\alpha)\right]^2}$$

The preliminary findings from Sect. 3.1 are now confirmed by the validation procedure, which clearly shows that:

$$Y^{\mathrm{PCK}}(\mathbf{x}_\alpha) \leq Z^{\mathrm{CK}}(\mathbf{x}_\alpha) \leq Y^{\mathrm{CK}}(\mathbf{x}_\alpha),$$

meaning that the iterative algorithm is successfull. Regarding the scores, if the correlation is a bit less with the new algorithm, its RMSE is better. As a consequence, not only the estimations satisfy the inequality constraint but they are also consistent with the observations.

From the cross-validation procedure described previously, we respectively store the values of $\#\{\mathcal{B}_\alpha\}$ and $\#\{\mathcal{B}_\beta\}$, the number of basic weights for the main and secondary variable. A high number of basic weights indicates that only a few number of weights are fixed in the iterative process so that the $PM_{2.5}$ estimations satisfy the physical inequality. On the contrary, extremely low values for the quantities means that most of the observations available are used with their non-basic weighting in the estimation, leading to a simple heuristic. Figure 5b displays the bivariate distribution of the basic weights. As we can see, only a few samples display low values for both $\#\{\mathcal{B}_\alpha\}$ and $\#\{\mathcal{B}_\beta\}$. Even when it is the case, and because we use definition (25) for the most non-basic weights identification, the nearest observations are always used to compute the $PM_{2.5}$ optimal interpolation, which explains why the estimations remain consistent. In terms of computation cost, let precise that because the number of modified weights are small, the penalized version cost is very close to the original cokriging algorithm. When the inequality constraint is not satisfied and the iterative penalized scheme is involved, less than 10 iterations were generally involved in our datasets (i.e. less than 10 observational weights are non-basic) and no estimation location ended with a simple heuristic interpolation (when the number of iteration reaches the number of observations, thus excluding most of the available information). This specific cases happened when $PM_{2.5}$ and PM_{10} observations are both close in terms of values and spatial locations.

4. Conclusions

The problem of consistency in cokriging arises when dealing with quantities that involve inequality constraints. In

a study by Beauchamp et al. [3], cokriging was employed to enhance the estimation of $PM_{2.5}$ by incorporating PM_{10} observations using a specific additive modeling approach. The local means were derived by exploiting the physical relationship between the two variables. Although this cokriging approach led to improved estimations, it failed to ensure that the resulting estimations adhered to the inequality constraint. From a mathematical standpoint, it is possible to verify this by conducting conditional simulations at a target location $\mathbf{x}_0$. Such simulations reveal that the intersection of the $PM_{2.5}$ simulated distribution with the PM_{10} simulated distribution is never empty. Thus, even when the average $PM_{2.5}$ simulation is higher than the average PM_{10} simulation, the consistency of the estimation is preserved.

To address the issue of inconsistencies in $PM_{2.5}$ concentrations, a new algorithm is proposed, in which additional constraints are introduced on the cokriging weights. They allow for successive iterations to solve the cokriging system in terms of basic component that satisfies these new constraints. The algorithm also includes several computational details that ensure a reasonable computational cost, making it practical for operational contexts.

Although the new estimator is performing well in most cases, there are still some limitations that need to be considered. One such limitation is the algorithm sensitivity to observation noise, as inaccuracies in the observed data can lead to significant errors in the final mapping. This can be especially problematic in areas with a high degree of spatial variability, where the data may be sparse or irregularly distributed. In such cases, the successive iterations in the penalization procedure may reduce the set of basic weights too far and thus build a simple heuristic that poorly estimates the true $PM_{2.5}$ concentration. This can result in a less accurate final mapping and may require additional adjustments to improve the algorithm's performance. Other limitations of the proposed algorithm include its reliance on assumptions about the underlying spatial structure of the data, as well as its potential limitations in handling non-Gaussian and non-stationary data, which could be crucial for communication and decision-making.

Finally, the use of our algorithm may apply to similar problems: it would remain identical and valid. First example would be the estimation of PM_non-volatile fraction, see e.g. [5]. Only the penalization $\overline{\lambda}$ and $\overline{\nu}$ introduced on the cokriging weights shall be adapted.

Acknowledgements
This study was partially funded by the French Ministry in charge of the environment, when the authors were members of the French Central Laboratory for Air Quality Monitoring (LCSQA).

Author contribution statement
Maxime Beauchamp: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper. **Bertrand Bessagnet**: Analyzed and interpreted the data; Wrote the paper.

Data availability statement
The software developed is now available through a DOI produced via Zenodo. The dataset comes from a public database (CAMS Copernicus) but since this work was made 4-5 years it was not possible to fully retrieve the exact dataset used at this time. We prefer to make the code public with associated data formatting so that any potential user can adapt and test the code on new datasets.

Declaration of interest's statement
The authors declare no conflict of interest.

References

[1] P. Abrahamsen and F. E. Benth. Kriging with inequality constraints. *Mathematical Geology*, 33(6):719–744, 2001. ISSN 1573-8868. doi: 10.1023/A:1011078716252. URL http://dx.doi.org/10.1023/A:1011078716252.

[2] R.J. Barnes and T.B. Johnson. Positive kriging. In G. Verly, M. David, A.G. Journel, and A. Marechal, editors, *Geostatistics for Natural Resources Characterization*, pages 231–244. Springer Netherlands, 1984. ISBN 978-94-010-8157-3. doi: 10.1007/978-94-009-3699-7_14. URL http://dx.doi.org/10.1007/978-94-009-3699-7_14.

[3] M. Beauchamp, B. Bessagnet, C. de Fouquet, L. Malherbe, F. Meleux, and A. Ung. An additive geostatistical model for mixing total and partial PM_{10} observations with CHIMERE rCTM. *Atmospheric Environment*, 189:61 – 79, 2018. ISSN 1352-2310. doi: https://doi.org/10.1016/j.atmosenv.2018.06.035. URL http://www.sciencedirect.com/science/article/pii/S1352231018304266.

[4] B. Bessagnet, E. Debry, and L. Malherbe. Bilan des mesures de PM_{10} et $PM_{2.5}$ ajustées et évaluation des outils de modélisation. Technical Report DRC-10-111613-05227A, LCSQA, 2010. URL www.lcsqa.org.

[5] Bertrand Bessagnet, Nadine Allemand, Jean-Philippe Putaud, Florian Couvidat, Jean-Marc André, David Simpson, Enrico Pisoni, Benjamin N Murphy, and Philippe Thunis. Emissions of carbonaceous particulate matter and ultrafine particles from Vehicles-A scientific review in a Cross-Cutting context of air pollution and climate change. *Appl Sci (Basel)*, 12(7):1–52, April 2022.

[6] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, 2004. ISBN 0-521-83378-7.

[7] P. J. Brown, N.D. Le, and J.V. Zidek. Multivariate spatial interpolation and exposure to air pollutants. *Canadian Journal of Statistics*, 22(4):489–509, 1994. ISSN 1708-945X. doi: 10.2307/3315406. URL http://dx.doi.org/10.2307/3315406.

[8] 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.

[9] J-P. Chilès, S. Séguret, and P-M. Riboud. Geostatistical analysis of validation data of an air pollution simulator. In *8th International geostatistics Congress*, page on CD, Santiago, Chile, December 2008. URL https://hal-mines-paristech.archives-ouvertes.fr/hal-00776934.

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

[11] CIESIN. Global annual average $PM_{2.5}$ grids from modis and misr aerosol optical depth (aod), 20160510 2013. URL http://dx.doi.org/10.7927/H4H41PB4.

[12] S. Conrad. Armadillo: an open source C++ linear algebra library for fast prototyping and computationally intensive experiments. Technical report, NICTA, 2010.

[13] N. Cressie and C. K. Wikl. The variance-based cross-variogram you can add apples and oranges. *Mathematical Geology*, 30:789–800, 1998.

[14] C. V. Deutsch. Correcting for negative weights in ordinary kriging. *Comput. Geosci.*, 22(7):765–773, August 1996. ISSN 0098-3004. doi: 10.1016/0098-3004(96)00005-2. URL http://dx.doi.org/10.1016/0098-3004(96)00005-2.

[15] Directive 2008/50/EC. *Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe*, 2008.

[16] O. Dubrule and C. Kostov. An interpolation method taking into account inequality constraints: I. methodology. *Mathematical Geology*, 18(1):33–51, 1986.

[17] D. Eddelbuettel and C. Sanderson. Rcpparmadillo: Accelerating R with high-performance C++ linear algebra. *Computational Statistics and Data Analysis*, 71:1054–1063, March 2014. URL http://dx.doi.org/10.1016/j.csda.2013.02.005.

[18] A.E. Gelfand, A.M. Schmidt, S. Banerjee, and C.F. Sirmans. Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, 13(2):263–312, Dec 2004. ISSN 1863-8260. doi: 10.1007/BF02595775. URL https://doi.org/10.1007/BF02595775.

[19] L. E. Gerharz and E. Pebesma. Using geostatistical simulation to disaggregate air quality model results for individual exposure estimation on gps tracks. *Stochastic Environmental Research and Risk Assessment*, 27(1):223–234, 2013. ISSN 1436-3259. doi: 10.1007/s00477-012-0578-9. URL http://dx.doi.org/10.1007/s00477-012-0578-9.

[20] G. B. M. Heuvelink, E. Pebesma, and B. Gräler. *Encyclopedia of GIS*, chapter Space-Time Geostatistics, pages 1–7. Springer International Publishing, Cham, 2015. ISBN 978-3-319-23519-6. doi: 10.1007/978-3-319-23519-6_1647-1. URL http://dx.doi.org/10.1007/978-3-319-23519-6_1647-1.

[21] E. H. Isaaks and R. M. Srivastava. *An Introduction to Applied Geostatistics*. Oxford University Press, New-York, 1989. ISBN 0-19-505012-6.

[22] K. Krivoruchko and A. Gribov. *geoENV IV — Geostatistics for Environmental Applications: Proceedings of the Fourth European Conference on Geostatistics for Environmental Applications held in Barcelona, Spain, November 27–29, 2002*, chapter Geostatistical Interpolation and Simulation in the Presence of Barriers, pages 331–342. Springer Netherlands, Dordrecht, 2004. ISBN 978-1-4020-2115-2. doi: 10.1007/1-4020-2115-1_28. URL http://dx.doi.org/10.1007/1-4020-2115-1_28.

[23] V. Langlais. Krigeage sous contraintes d'inégalités en voisinage glissant. Technical Report Note interne, Centre de Géostatistique, ENSMP, Fontainebleau, 1988.

[24] S. Mailler, L. Menut, D. Khvorostyanov, M. Valari, F. Couvidat, G. Siour, S. Turquety, R. Briant, P. Tuccella, B. Bessagnet, A. Colette, L. Létinois, K. Markakis, and F. Meleux. Chimere-2017: from urban to hemispheric chemistry-transport modeling. *Geoscientific Model Development*, 10(6):2397–2423, 2017. doi: 10.5194/gmd-10-2397-2017. URL https://www.geosci-model-dev.net/10/2397/2017/.

[25] S. Mailler, L. Menut, D. Khvorostyanov, M. Valari, F. Couvidat, G. Siour, S. Turquety, R. Briant, P. Tuccella, B. Bessagnet, A. Colette, L. Létinois, K. Markakis, and F. Meleux. Chimere-2017: from urban to hemispheric chemistry-transport modeling. *Geoscientific Model Development*, 10(6):2397–2423, 2017. doi: 10.5194/gmd-10-2397-2017. URL https://www.geosci-model-dev.net/10/2397/2017/.

[26] L Malherbe and A. Ung. Travaux relatifs á la plate-forme nationale de modélisation PREV'AIR : Réalisation de cartes analysées d'Ozone. Technical Report DRC-10-103351-01139A, LCSQA, 2009. URL www.lcsqa.org.

[27] A. M. Michalak. A gibbs sampler for inequality-constrained geostatistical interpolation and inverse modeling. *Water Resources Research*, 44(9):n/a–n/a, 2008. ISSN 1944-7973. doi: 10.1029/2007WR006645. URL http://dx.doi.org/10.1029/2007WR006645. W09437.

[28] A. M. Michalak and P. K. Kitanidis. A method for enforcing parameter nonnegativity in bayesian inverse problems with an application to contaminant source identification. *Water Resources Research*, 39 (2):n/a–n/a, 2003. ISSN 1944-7973. doi: 10.1029/2002WR001480. URL http://dx.doi.org/10.1029/2002WR001480. 1033.

[29] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL https://www.R-project.org/.

[30] M. C. Ribeiro, P. Pinho, C. Branquinho, E. Llop, and M. J. Pereira. Geostatistical uncertainty of assessing air quality using high-spatial-resolution lichen data: A health study in the urban area of sines, portugal. *Science of The Total Environment*, 562:740 – 750, 2016. ISSN 0048-9697. doi: http://dx.doi.org/10.1016/j.scitotenv.2016.04.081. URL http://www.sciencedirect.com/science/article/pii/

S0048969716307616.

[31] J. Rivoirard and T. Romary. Continuity for kriging with moving neighborhood. *Mathematical Geosciences*, 43(4):469–481, 2011. ISSN 1874-8953. doi: 10.1007/s11004-011-9330-0. URL http://dx.doi.org/10.1007/s11004-011-9330-0.

[32] R. I. Rothenberg. *Linear Programming*. Oxford, North-Holland, New York, 1979.

[33] L. Rouil, C. Honoré, R. Vautard, M. Beekmann, B. Bessagnet, L. Malherbe, F. Meleux, A. Dufour, C. Elichegaray, J.-M. Flaud, L. Menut, D. Martin, A. Peuch, V.-H. Peuch, and N. Poisson. Prev'air: An operational forecasting and mapping system for air quality in europe. *Bulletin of the American Meteorological Society*, 90(1):73–84, 2009. doi: 10.1175/2008BAMS2390.1. URL https://doi.org/10.1175/2008BAMS2390.1.

[34] P.D. Sampson, M. Richards, A. A. Szpiro, S. Bergen, L. Sheppard, T. V. Larson, and J. D. Kaufman. A regionalized national universal kriging model using partial least squares regression for estimating annual $PM_{2.5}$ concentrations in epidemiology. *Atmospheric Environment*, 75:383 – 392, 2013. ISSN 1352-2310. doi: http://dx.doi.org/10.1016/j.atmosenv.2013.04.015. URL http://www.sciencedirect.com/science/article/pii/S1352231013002604.

[35] A.M. Schmidt and A.E. Gelfand. A bayesian coregionalization approach for multivariate pollutant data. *Journal of Geophysical Research: Atmospheres*, 108(D24), 2003. ISSN 2156-2202. doi: 10.1029/2002JD002905. URL http://dx.doi.org/10.1029/2002JD002905. 8783.

[36] S. Soltani-Mohammadi and A. E. Tercan. Constrained multiple indicator kriging using sequential quadratic programming. *Computers & Geosciences*, 48:211 – 219, 2012. ISSN 0098-3004. doi: https://doi.org/10.1016/j.cageo.2012.01.003. URL http://www.sciencedirect.com/science/article/pii/S0098300412000052.

[37] F. Szidarovszky, E. Y. Baafi, and Y. C. Kim. Kriging without negative weights. *Mathematical Geology*, 19(6):549–559, Aug 1987. ISSN 1573-8868. doi: 10.1007/BF00896920. URL https://doi.org/10.1007/BF00896920.

[38] H. Wackernagel. *Multivariate Geostatistics. An Introduction with Applications*. Springer-Verlag Berlin Heidelberg, New York, 2003. doi: 10.1007/978-3-662-05294-5. ISBN 978-3-540-44142-7.

[39] Y. Yang, G. Christakos, W. Huang, C. Lin, P. Fu, and Y. Mei. Uncertainty assessment of $PM_{2.5}$ contamination mapping using spatiotemporal sequential indicator simulations and multi-temporal monitoring data. *Scientific Reports*, 6:24335 EP –, Apr 2016. URL http://dx.doi.org/10.1038/srep24335. Article.

Appendices

A. Variance of the penalized cokriging estimator

The related cokriging variance of $Y^{\mathrm{PCK}}(\mathbf{x}_0)$ is:

$$
\begin{aligned}
&\mathrm{Var}\left[Y(\mathbf{x}_0) - Y^{\mathrm{PCK}}(\mathbf{x}_0)\right] = \sigma^2_{\mathrm{PCK}}(\mathbf{x}_0) \\
&= C^Y(\mathbf{0}) + \left(\sum_{\substack{\alpha\in\mathcal{B}_\alpha \\ \alpha'\in\mathcal{B}_\alpha}} \lambda_\alpha\lambda_{\alpha'} C^1_{\alpha\alpha'} + \sum_{\substack{\alpha\in\overline{\mathcal{B}_\alpha} \\ \alpha'\in\overline{\mathcal{B}_\alpha}}} \lambda_\alpha\lambda_{\alpha'} C^1_{\alpha\alpha'} \right. \\
&\quad \left. + \sum_{\substack{\alpha\in\mathcal{B}_\alpha \\ \alpha'\in\overline{\mathcal{B}_\alpha}}} \lambda_\alpha\lambda_{\alpha'} C^1_{\alpha\alpha'} + \sum_{\substack{\alpha\in\overline{\mathcal{B}_\alpha} \\ \alpha'\in\mathcal{B}_\alpha}} \lambda_\alpha\lambda_{\alpha'} C^1_{\alpha\alpha'} \right) \\
&\quad + \left(\sum_{\substack{\beta\in\mathcal{B}_\beta \\ \beta'\in\mathcal{B}_\beta}} \nu_\beta\nu_{\beta'} C^2_{\beta\beta'} + \sum_{\substack{\beta\in\overline{\mathcal{B}_\beta} \\ \beta'\in\overline{\mathcal{B}_\beta}}} \nu_\beta\nu_{\beta'} C^2_{\beta\beta'} \right. \\
&\quad \left. + \sum_{\substack{\beta\in\mathcal{B}_\beta \\ \beta'\in\overline{\mathcal{B}_\beta}}} \nu_\beta\nu_{\beta'} C^2_{\beta\beta'} + \sum_{\substack{\beta\in\overline{\mathcal{B}_\beta} \\ \beta'\in\mathcal{B}_\beta}} \nu_\beta\nu_{\beta'} C^2_{\beta\beta'} \right) \\
&\quad - 2\left(\sum_{\alpha\in\mathcal{B}_\alpha} \lambda_\alpha C^1_{\alpha 0} + \sum_{\alpha\in\overline{\mathcal{B}_\alpha}} \lambda_\alpha C^1_{\alpha 0} \right. \\
&\quad \left. + \sum_{\beta\in\mathcal{B}_\beta} \nu_\beta C^{12}_{\beta 0} + \sum_{\beta\in\overline{\mathcal{B}_\beta}} \nu_\beta C^{12}_{\beta 0} \right) \\
&\quad + 2\left(\sum_{\substack{\alpha\in\mathcal{B}_\alpha \\ \beta\in\mathcal{B}_\beta}} \lambda_\alpha\nu_\beta C^{12}_{\alpha\beta} + \sum_{\substack{\alpha\in\overline{\mathcal{B}_\alpha} \\ \beta\in\mathcal{B}_\beta}} \lambda_\alpha\nu_\beta C^{12}_{\alpha\beta} \right. \\
&\quad \left. + \sum_{\substack{\alpha\in\mathcal{B}_\alpha \\ \beta\in\overline{\mathcal{B}_\beta}}} \lambda_\alpha\nu_\beta C^{12}_{\alpha\beta} + \sum_{\substack{\alpha\in\overline{\mathcal{B}_\alpha} \\ \beta\in\overline{\mathcal{B}_\beta}}} \lambda_\alpha\nu_\beta C^{12}_{\alpha\beta} \right)
\end{aligned}
$$

By regrouping all the terms from Eq. (19a) to (19d), this

kriging variance is:

$$\begin{aligned}
&\sigma^2_{\mathrm{PCK}}(\mathbf{x}_0)\\
&= C^Y(\mathbf{0})\\
&- \sum_{\alpha\in\mathcal{B}_\alpha} \lambda_\alpha C^1_{\alpha 0} - (\mu_0 + \sum_i \upsilon_i f^i(\mathbf{x}_\alpha))\\
&- \sum_{\alpha\in\overline{\mathcal{B}_\alpha}} \lambda_\alpha \left[C^1_{\alpha 0} - \tau_\alpha\right] - (\mu_0 + \sum_i \upsilon_i f^i(\mathbf{x}_\alpha))\\
&- \sum_{\beta\in\mathcal{B}_\beta} \nu_\beta C^{12}_{\beta 0} - (\upsilon_0 + \sum_i \upsilon_i f^i(\mathbf{x}_\beta) + \sum_j \mu_j g^j(\mathbf{x}_\beta))\\
&- \sum_{\beta\in\overline{\mathcal{B}_\beta}} \nu_\beta \left[C^{12}_{\beta 0} - \omega_\beta\right] - (\upsilon_0 + \sum_i \upsilon_i f^i(\mathbf{x}_\beta)\\
&+ \sum_j \mu_j g^j(\mathbf{x}_\beta))
\end{aligned}$$

finally leading to:

$$\begin{aligned}
\sigma^2_{\mathrm{PCK}}(\mathbf{x}_0) = C^1(\mathbf{0}) &- \sum_{\alpha\in\mathcal{B}_\alpha} \lambda_\alpha C^1_{\alpha 0} - \sum_{\alpha\in\overline{\mathcal{B}_\alpha}} \lambda_\alpha \left[C^1_{\alpha 0} - \tau_\alpha\right]\\
&- \sum_{\beta\in\mathcal{B}_\beta} \nu_\beta C^{12}_{\beta 0} - \sum_{\beta\in\overline{\mathcal{B}_\beta}} \nu_\beta \left[C^{12}_{\beta 0} - \omega_\beta\right]\\
&- \mu_0 - \sum_i \upsilon_i f^i(\mathbf{x}_0)
\end{aligned}$$

B. Continuous cokriging for moving neighborhood

In the usual cokriging framework, a moving neighborhood is more costly than a unique neighborhood because the kriging matrix has to be inverted for each target location $\mathbf{x}_0$. In our penalized algorithm, because the system is updated at each step of the algorithm, the smaller this linear system is, the faster is the update. In addition, a moving neighborhood allows to refine at best the coefficients of the drift according to the local behavior of the covariates to the observations. Still, one issue of using a moving neighborhood is than it can create artificial discontinuities in the estimation when mapping the field on a (regular) grid.

In Rivoirard and Romary [31], the discontinuities caused by the moving neighbourhood are managed by considering the observations $Z(\mathbf{x}_\beta)$ spoiled by a noise $\varepsilon_Z(\mathbf{x}_\beta)$.
As a consequence, the kriging matrix is modified: to each term $\mathrm{Cov}\left[Z(\mathbf{x}_\beta), Z(\mathbf{x}_{\beta'})\right]$ is added $\mathrm{Cov}\left[\varepsilon_Z(\mathbf{x}_\beta), \varepsilon_Z(\mathbf{x}_{\beta'})\right]$. The same applies for $\mathrm{Cov}\left[Y(\mathbf{x}_\alpha), Y(\mathbf{x}_{\alpha'})\right]$ and the kriging fashion is easily transposed for multivariate datasets. In particular, the cokriging becomes:

$$\begin{aligned}
Y^{\mathrm{CK}}(\mathbf{x}_0) &= \sum_\alpha \lambda_\alpha \{Y(\mathbf{x}_\alpha) + \varepsilon_Y(\mathbf{x}_\alpha)\}\\
&+ \sum_\beta \nu_\beta \{Z(\mathbf{x}_\beta) + \varepsilon_Z(\mathbf{x}_\beta)\}
\end{aligned}$$

When the initial covariance structure is non-continuous (with a nugget effect), $\varepsilon_Z(\mathbf{x}_\beta)$ and $\varepsilon_Z(\mathbf{x}_{\beta'})$ are considered independent when $\beta \neq \beta'$ and $\mathrm{Cov}\left[\varepsilon_Z(\mathbf{x}_\beta), \varepsilon_Z(\mathbf{x}_{\beta'})\right] = 0$. The same applies for ε_Y. The variances $\mathrm{Var}\left[\varepsilon_Z(\mathbf{x}_\alpha)\right]$ and $\mathrm{Var}\left[\varepsilon_Y(\mathbf{x}_\beta)\right]$ of the noises ε_Z and ε_Y are chosen by the user: they increases according to the distance to $\mathbf{x}_0$ and are neglected or even set to 0 for the nearest data points:

$$\begin{aligned}
\mathrm{Var}\left[\varepsilon_Z(\mathbf{x}_\alpha)\right] &= C_Z(0)\left(\frac{\mathbf{h}_{\alpha 0} - r}{R - \mathbf{h}_{\alpha 0}}\right)^2\\
\mathrm{Var}\left[\varepsilon_Y(\mathbf{x}_\beta)\right] &= C_Y(0)\left(\frac{\mathbf{h}_{\beta 0} - r}{R - \mathbf{h}_{\beta 0}}\right)^2
\end{aligned}$$

with $\mathbf{h}_{\alpha 0} = ||\mathbf{x}_\alpha - \mathbf{x}_0||$, $\mathbf{h}_{\beta 0} = ||\mathbf{x}_\beta - \mathbf{x}_0||$, R the radius of the moving neighborhood and $r = p.R$, with $p < 1$.

The lagrangian L is now defined as:

$$\begin{aligned}
L(\mathbf{x}_0) &= \mathrm{Var}\left[Y(\mathbf{x}_0) - Y^{\mathrm{CK}}(\mathbf{x}_0)\right] + 2\mu_0\left(\sum_\alpha \lambda_\alpha - 1\right)\\
&+ 2\upsilon_0 \sum_\beta \nu_\beta + 2\mu_j\left(\sum_\alpha \lambda_\alpha g^j(\mathbf{x}_\alpha) - g^j_0\right)\\
&+ 2\upsilon_i\left(\sum_\beta \nu_\beta f^i(\mathbf{x}_\beta) - f^i_0\right) + \sum_\alpha \lambda^2_\alpha n_\alpha + \sum_\beta \nu^2_\beta n_\beta
\end{aligned}$$

where the weights λ_α and ν_β are respectively penalized by the quantities $n_\alpha = \mathrm{Var}\left[\varepsilon_Y(\mathbf{x}_\beta)\right]$ and $n_\beta = \mathrm{Var}\left[\varepsilon_Z(\mathbf{x}_\alpha)\right]$.

Equating the partial derivatives $\frac{\partial L}{\partial \lambda(\mathbf{x}_\alpha)}$ and $\frac{\partial L}{\partial \nu(\mathbf{x}_\beta)}$ to zero leads to:

$$\begin{aligned}
&\lambda_\alpha n_\alpha + \sum_{\alpha'} \lambda_{\alpha'} C^Y_{\alpha\alpha'} - C^Y_{\alpha 0} + \sum_\beta \nu_\beta C^{12}_{\alpha\beta}\\
&+ \mu_0 + \upsilon_i f^i(\mathbf{x}_\alpha) + \mu_j g^j(\mathbf{x}_\alpha) &&= 0
\end{aligned}$$

and

$$\begin{aligned}
&\nu_\beta n_\beta + \sum_{\beta'} \nu_{\beta'} C^2_{\beta\beta'} - C^{12}_{\beta 0} + \sum_\alpha \lambda_\alpha C^{12}_{\alpha\beta}\\
&+ \upsilon_0 + \upsilon_i f^i(\mathbf{x}_\beta) &&= 0
\end{aligned}$$

and the cokriging matrix system is:

$$\begin{aligned}
&(\mathbf{K} + \mathbf{N})\mathbf{W} + \mathbf{A}\mathbf{M} &&= \mathbf{K_0}\\
&\mathbf{A}'\mathbf{W} &&= \mathbf{F_0}
\end{aligned}$$

where $\mathbf{N} = \mathbf{I}\mathbf{n}'$ and $\mathbf{n} = \begin{bmatrix}\mathbf{n}_\alpha\\ \mathbf{n}_\beta\end{bmatrix}_{N+M}$.

Finally, the conditions (19a) to (19d) related to the estimator $Y^{\mathrm{PCK}}(\mathbf{x}_0)$ can thus be rewritten as follows, if the estimation is made with a moving neighbourhood in which the

discontinuities vanishes thanks to the noises n_α and n_β:

$$
\begin{aligned}
&(\mathbf{K}_{11}^{\mathrm{bb}} + \mathbf{n}_{1\mathrm{b}})\lambda_{\mathrm{b}} + \mathbf{K}_{12}^{\mathrm{bb}}\nu_{\mathrm{b}} + \mu_0 + \boldsymbol{\mu}_j \mathbf{g}_{\mathrm{b}}^j \\
&= \mathbf{K}_{10}^{\mathrm{b}} - (\mathbf{K}_{11}^{\mathrm{bn}} + \mathbf{n}_{1\mathrm{n}})\overline{\lambda_{\mathrm{n}}} - \mathbf{K}_{12}^{\mathrm{bn}}\overline{\nu_{\mathrm{n}}} \\
&(\mathbf{K}_{11}^{\mathrm{nb}} + \mathbf{n}_{1\mathrm{n}})\lambda_{\mathrm{b}} + \mathbf{K}_{12}^{\mathrm{nb}}\nu_{\mathrm{b}} + \mu_0 + \boldsymbol{\mu}_j \mathbf{g}_{\mathrm{n}}^j + \mathbf{1}_{11}^{\mathrm{nn}}\boldsymbol{\tau}_{\mathrm{n}} \\
&= \mathbf{K}_{10}^{\mathrm{n}} - (\mathbf{K}_{11}^{\mathrm{nn}}\overline{\lambda_{\mathrm{n}}} + \mathbf{n}_{1\mathrm{n}}) - \mathbf{K}_{12}^{\mathrm{nn}}\overline{\nu_{1\mathrm{n}}} \\
&\mathbf{K}_{21}^{\mathrm{bb}}\lambda_{\mathrm{b}} + (\mathbf{K}_{22}^{\mathrm{bb}}\nu_{\mathrm{b}} + \mathbf{n}_{2\mathrm{b}}) + \upsilon_0 + \boldsymbol{\upsilon}_i \mathbf{f}_{\mathrm{b}}^j \\
&= \mathbf{K}_{20}^{\mathrm{b}} - \mathbf{K}_{21}^{\mathrm{bn}}\overline{\lambda_{\mathrm{n}}} - (\mathbf{K}_{22}^{\mathrm{bn}} + \mathbf{n}_{2\mathrm{n}})\overline{\nu_{\mathrm{n}}} \\
&\mathbf{K}_{21}^{\mathrm{nb}}\lambda_{\mathrm{b}} + (\mathbf{K}_{22}^{\mathrm{nb}} + \mathbf{n}_{2\mathrm{n}})\nu_{\mathrm{b}} + \upsilon_0 + \boldsymbol{\upsilon}_i \mathbf{f}_{\mathrm{n}}^j + \mathbf{1}_{22}^{\mathrm{nn}}\boldsymbol{\omega}_{\mathrm{n}} \\
&= \mathbf{K}_{20}^{\mathrm{n}} - \mathbf{K}_{21}^{\mathrm{nn}}\overline{\lambda_{\mathrm{n}}} - (\mathbf{K}_{22}^{\mathrm{nn}} + \mathbf{n}_{2\mathrm{n}})\overline{\nu_{\mathrm{n}}}
\end{aligned}
$$

$\mathbf{n}_{1\mathrm{b}}$ is the vector of noises $\{n(\mathbf{x}_\alpha)\},\ \alpha \in \mathcal{B}_\alpha$.
$\mathbf{n}_{1\mathrm{n}}$ is the vector of noises $\{n(\mathbf{x}_\alpha)\},\ \alpha \in \overline{\mathcal{B}_\alpha}$.
$\mathbf{n}_{2\mathrm{b}}$ is the vector of noises $\{n(\mathbf{x}_\beta)\},\ \beta \in \mathcal{B}_\beta$.
$\mathbf{n}_{2\mathrm{n}}$ is the vector of noises $\{n(\mathbf{x}_\beta)\},\ \beta \in \overline{\mathcal{B}_\beta}$.

Let us note that in air quality kriging-based maps, the discontinuities mainly arises on very large domain of estimations, over Europe for instance, where the monitoring network is not homogeneously distributed: a lot of data are available in Western and Central Europe while the network is sparse elsewhere. Thus, to enable a local fitting of the drift, a number-based neighborhood $\delta_{\alpha 0} = \{\alpha, ||\mathbf{x}_\alpha - \mathbf{x}_0|| < ||\mathbf{x}_N - \mathbf{x}_0||\}$, where $\mathbf{x}_N$ denotes the Nth nearest neighbour of $\mathbf{x}_0$, is prefered to a distance-based neighborhood $\delta_{\alpha 0} = \{\alpha, ||\mathbf{x}_\alpha - \mathbf{x}_0|| \leq D\}$; the latter including too many stations where the network is dense, and too few in the badly informed areas. In Rivoirard and Romary [31], the continuous kriging is distance-based driven. To overcome this problem, and because the estimations are done on regular grids, the distance $R(\mathbf{x}_0)$ between the target location $\mathbf{x}_0$ and the Nth nearest neighbor of $\mathbf{x}_0$ is computed for each gridcell. This distance spatially varies but in a continuous way since the grid is regular (see Fig. B.6). As a consequence, the distance-based neighborhood approach is kept but the radius used is no longer spatially constant: $R(\mathbf{x}_0)$ is substituted to R in Eq.(26).

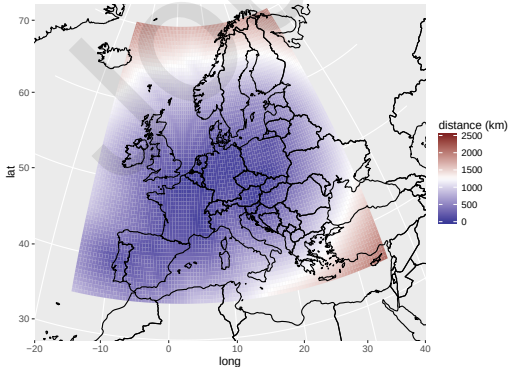


Figure B.6: Distance between the gricells $\mathbf{x}_0$ and the 20th nearest neighbour

Declaration of interests

X The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: