



HAL
open science

Mixture Kriging on granular data

Marc Grossouvre, Didier Rullière

► **To cite this version:**

| Marc Grossouvre, Didier Rullière. Mixture Kriging on granular data. 2022. hal-03276127v2

HAL Id: hal-03276127

<https://hal.science/hal-03276127v2>

Preprint submitted on 25 Aug 2022 (v2), last revised 17 Jan 2024 (v4)

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

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

Mixture Kriging on granular data

Marc Grossouvre*

Didier Rullière†

Wednesday 24th August, 2022‡

Abstract

This paper deals with three related problems in a geostatistical context. First, some data are available for given areas of the space rather than for some point locations which creates problems of multiscale areal data. Second, some uncertainties rely both on the input locations and on measured quantities at these locations, involving uncertainty propagation problems. Third, multidimensional outputs can be observed, with sometimes missing data. These three problems are addressed simultaneously here by considering mixtures of multivariate random fields and by adapting standard Kriging methodology to this context. While the usual Gaussian setting is lost, we show that conditional mean, variance and covariance can be derived from this specific setting. Case studies are presented both with simulated data and real data. In particular, we discuss the question of information loss in learning buildings energy efficiency.

Keywords— Mixture Kriging, granular data, multiscale processes, regional Kriging, area-to-point, areal data, block Kriging, change of support, ecological inference, disaggregation.

1 Introduction

Spatial interpolation In the geostatistical field, spatial interpolation aims at predicting the value of an output $y(x) \in \mathbb{R}$ that depends on an input $x \in \chi$. The input x can be a geographical location ($\chi = \mathbb{R}^2$), or for instance an industrial design summarized by some parameters $\chi = \mathbb{R}^d$, $d \in \mathbb{N}^*$). The output $y(x)$ can be any quantity of interest (pollution level, performance of a design, etc.), a real value in the simplest case.

The output $y(x)$ is not necessarily observed for all possible values of x in χ , because of lack of measurement devices, of the financial cost of a physical experiment, or because of the time required to run a computer experiment when $y(\cdot)$ is issued from computer simulation, as is the case for meteorological general circulation models (GCM, see Lima et al. [2021]). The purpose is thus to infer the value of $y(x)$ for some value of x , given some observations $\{(x_i, y(x_i))\}_{i=1, \dots, n}$. Many spatial interpolation techniques are available (splines, inverse distance weighting, nearest neighbours, regression models, etc.). Among these popular techniques, the Gaussian Process Regression is very well known; it is also known as Kriging, with an optional release of the underlying Gaussian assumption. It models the different possible values of $y(x)$ by a conditional Gaussian process $Y(x)$ given $Y(x_i) = y(x_i)$, $i = 1, \dots, n$. In the simplest case where the unconditional (prior) process $Y(\cdot)$ is a Gaussian process with mean zero, the simple Kriging predictor ends up in a (posterior) conditional Gaussian process having conditional Kriging mean and covariance, for any $x, x' \in \chi$:

$$\begin{cases} \mathbb{E}[Y(x)|Y(x_i) = y(x_i), i = 1, \dots, n] & = \mathbf{h}(x)^\top \mathbf{K}^{-1} \mathbf{y} \\ \text{Cov}[Y(x), Y(x')|Y(x_i) = y(x_i), i = 1, \dots, n] & = k(x, x') - \mathbf{h}(x)^\top \mathbf{K}^{-1} \mathbf{h}(x') \end{cases} \quad (1)$$

*Marc Grossouvre, U.R.B.S. SAS, Bâtiment des Hautes Technologie, 20 Rue Professeur Benoît LAURAS, 42000 Saint-Etienne, France, marcgrossouvre@urbs.fr, website www.urbs.fr.

†Didier Rullière, Mines Saint-Etienne, Univ Clermont Auvergne, CNRS, UMR 6158 LIMOS, Institut Henri Fayol, Département GMI, Espace Fauriel, 29 rue Ponchardier, F - 42023 Saint-Etienne, France. didier.rulliere@emse.fr

‡version of this document: Wednesday 24th August, 2022, 16:26.

where $\mathbf{y} = (Y(x_i))_{i=1,\dots,n} \in \mathbb{R}^{n \times 1}$ is an observed column vector, and where all prior covariances at point level are assumed to be known:

$$\begin{cases} k(x, x') & := \text{Cov}[Y(x), Y(x')] \in \mathbb{R} \\ \mathbf{h}(x) & := (\text{Cov}[Y(x_i), Y(x)])_{i=1,\dots,n} \in \mathbb{R}^{n \times 1} \\ \mathbf{K} & := (\text{Cov}[Y(x_i), Y(x_j)])_{i,j \in \{1,\dots,n\}} \in \mathbb{R}^{n \times n} \end{cases}$$

and where \mathbf{K} is assumed to be invertible.

We aim here at proposing a spatial prediction technique when some data are available for given areas of the space, rather than for some specific locations. We also aim at handling uncertainties or noises relying on both input locations x and outputs $Y(x)$. At last we aim at considering multiple outputs, where some components of the observed outputs can be missing. To achieve this goal, we extend classical Kriging techniques to our new settings.

Literature There exists a vast literature dealing with predictions using areal data. Gotway and Young [2002] have published a review of methods under the clever and paradoxical title “Combining Incompatible Data”. The purpose of this review is to study an information that comes from a random field (either continuous or discrete) but that is available at a block (i.e. regional) level only. Computing a block value from the point field values is named aggregating data at block level. In general, additive variables such as population are excluded, and the focus is put on variables such as densities (e.g. concentration in mining context) for continuous fields or individual measurements (e.g. the height of a tree). Behind the seemingly diversity of methods, there is a constant approach: the observed output over an areal unit is an average of random variables, so that it is assumed that aggregation means averaging values. Average has a reduced dispersion as compared to point values. This is seen as positive but authors also note that it induces some problems (see details in Subsection 2.2).

A major field of development for addressing change of support problem with Kriging, beside mining, has been meteorological studies. Gelfand et al. [2001] have proposed a fully Bayesian approach to study ozone (measured at point level) and its effects (measured at ZIP level). Although this model is quite heavy to implement, we shall notice that the author rules out the mainstream approach that was to study the variogram “approximating” a block by its centroid. Moreover, in their model, the considered multivariate output is a time series.

Mathematically, a milestone has been set by in Kyriakidis [2004] with a complete Kriging model including area-to-point and sketching area-to-area prediction. This work has been cited a large number of times. Following that work, Goovaerts [2008] studies in depth the problem of estimating a variogram. He shows that averaging reduces the sill of the variogram and tries to tackle this bias. Those results have been successfully used by Poggio and Gimona [2015] for downscaling climate models and predicting soil wetness. In the latter work, Kriging is applied on residuals of a generalized additive model (see Wood [2017]) with a significant added value. Many other papers deal with using area-to-point Kriging or disaggregation (see Kerry et al. [2013], Truong and Heuvelink [2013], Yoo and Kyriakidis [2006]), area-to-area Kriging or block Kriging (see Zhang et al. [2018]), downscaling (see Jin et al. [2018], Pereira et al. [2018]).

Let us now detail identified problems in the literature, regarding the averaging approach. As stated before, a major impact of averaging random variables is the variance shrinkage of the average compared to the initial random variables that are averaged. An abundant research literature is available that tries to tackle the different problems generated by the systematic averaging in areal Kriging models. Let us quote first the Modifiable areal unit problem (MAUP, see the latest findings on this topic in Briz-Redón [2022]). It appears in situations such as studying crop yields: from one year to the other, the set of agricultural fields to aggregate for a certain type of crop varies. However, the correlations between output variables is strongly dependent on the aggregation process, making it impossible to compare correlations between two different years. We can also mention the ecological bias inference problem: correlations at individual level are different from correlations of averaged outputs at ecological (group level), lack of relevant information about individuals position leads to bias when studying averaged information about individuals distributed into areal units cross-classified by other individual (point level) variables (sex, race). And according to Gotway and Young: “The smoothing effect that results from averaging is the underlying cause of both the scale problem in the MAUP and aggregation bias in ecological studies.” [Gotway and Young, 2002]. The specific issue of variance reduction at block level has been compensated to some extent in Li et al. [2009] where a new method for covariance estimation is also proposed without integrating over a surface.

Despite those limitations, it is important to recall that averaging has turned out to be quite powerful for data suited to this kind of approach. In particular, satellite imaging has notably benefited from it. An illustrative example of this is the pan-sharpening process that is “a technique to combine the fine spatial resolution panchromatic (PAN) band with the coarse spatial resolution multispectral bands of the same satellite to create a fine spatial resolution multispectral image” Wang et al. [2016]. In this particular case, averaging is weighted over the satellite pixel: weighting for the average computation is made according to the distance with the centroid of the pixel. As in most articles we have consulted, aggregation is seen as a weighted integral but these weights are not regarded as a law of probability on the block, therefore ignoring some possible related statistics, and ignoring another potential source of stochastic dependence between blocks.

It is interesting to note that in the vast literature discussing change of support problems, blocks (also called areas) and points are considered as intrinsically different objects. The classification of problems such as “area-to-point” or “area-to-area” expresses well this categorization. The typology of problems that have been studied whether political Gotway and Young [2002] or medical Goovaerts [2006] for instance seems to assume that data is available either at point level or at aggregated level, not both. However, in most cases, there is no major difficulty in considering blocks made of a single point. Even if the distinction is necessary at some point such as inferring a point support covariance, we propose in this paper a model where learning and predicting can be from both aggregated and point support data. The notion of *grain* is introduced to express this new approach. This is consistent with research realities where one would like to be able to complement large aggregated open datasets with local observations. Similarly, blocks are usually viewed as connected surface areas in \mathbb{R}^2 that need to be “discretized” (Goovaerts [2008]) in order to make computations. But in many cases, actually in many problems, this strongly distorts reality where population for instance is clearly a set of discrete points heterogeneously located on a block (county, census tract...). The following concept of granular data makes no difference in nature between a grain containing a continuous set of points or a discrete one.

Granular data Let us consider an input space over which is defined a field of multidimensional random output variables. The specificity we introduce is that outputs may be defined and observed both for points of the input space and for some regions of this same input space. For instance, imagine that some sociological variables (salaries, expenses, etc.) are available for different geographical areas: cities, regions, countries, etc. Later on, these areas will be called grains, and we will develop a specific Kriging technique to handle this data. In this context, we refer to **granular data** and **grains** for these areas of the input space. For such data, one is interested in defining a suitable data model that is able to predict output variables for new inputs, be it points or grains. The underlying assumption in this work is that there is some dependence between outputs based on the relative positions of the associated inputs. Contrary to previous works presented in the litterature, we do not make any difference between observing a grain or a point (no averaging), nor do we approximate a grain with its centroid. We do not make any assumption on the grains shapes so much so that they can even be overlapping, partially or totally. And we do not make any difference between making predictions on grains or points.

A possible application of this model is in the field of geographic information, for instance to handle data that is released in open format by public or private institutions. Say for instance that a government releases the distribution of inhabitants salaries at municipality level. A private company may try to use this data to estimate the distribution of salaries at a smaller scale, say for a district in a city. And more than that, this company may include in its model both this institutional data and some known salaries at specific locations of the target territory. To handle this problem, we expose here a general Kriging approach that generalizes the usual Simple or Ordinary (Co)Kriging techniques. This kind of granular data is usual for satellite images for instance, in which case considered areas (grains) are very regular. But data is often available at levels of details that are very irregular. For instance, in geographical field, suppose that one has data for each municipality/county. It is a granularity with grains of very different shapes and sizes. They might even not be connected sets.

In the rest of the paper, we develop a model for the outputs that are available for different regions of the input space. These different regions form what we call here a **granularity**.

Specific problems This situation of data being partially available at different granularities is related to the following mathematical problems.

(P1) First, this is a multiscale change of support problem. Some outputs are not available for points

in the input space but only for some sets of input points (grains of inputs) of varying shapes and sizes. One thus needs to create a model to describe the link between the output associated with a single input (as usual random field model), and the output associated with a set of inputs. E.g. if a salary is given for a given country, does this quantity represent an average or something else? How is it connected with a salary at a specific site of the country?

- (P2) Second, there is an uncertainty propagation problem. The specificity of this kind of data creates an underlying uncertainty on the output variables at a given point location, because the knowledge at a larger scale derives from an aggregation of multiple locations: uncertainties depend both on the input locations and on measured quantities at these locations, which propagates into prediction uncertainty. Furthermore, modeling the uncertainty is also required to avoid some specific convexity bias: a function evaluated at a random location has different mean and variance than a function evaluated at an averaged location. E.g. if a salary is given for a given country, surely this salary must be considered random at a specific site, and resulting studies over a region should handle carefully this randomness: applying a convex function on an average salary will not only result in a loss of randomness, but also in a convexity bias as discussed further.
- (P3) Third, there may be missing data. Multidimensional outputs can be incompletely observed: not all the components of the outputs are observed. Hence a model able to handle multidimensional outputs with missing components is needed. E.g. one can imagine having salary, life expectancy, average taxes by cities, with life expectancy and/or salaries missing for some cities. We are most interested in predicting these missing values. And in general, we are interested in predicting any value of the output on any grain of our territory, even if this grain is restricted to a singleton. In this matter, we want to avoid the trap of completing incomplete outputs with averaged values, thus ignoring the data dispersion and possibly causing convexity bias as described in problem (P2).

What is done The originality of this paper is that it considers **mixture random variables** rather than averaged random variables over areas. One can note that Kriging has been used successfully with variables that are mixtures at point level (see for instance Lin et al. [2010]), but in our case, we make no specific assumption on the distribution at point level. We build the aggregated information over areas as mixtures. Averaging a large number of random variables lead to a reduction of the resulting variance, whereas mixing a large number of random variable does not tend to reduce the variance. As a consequence, the larger the considered area in our approach, the higher the uncertainty. We will show that this approach is well suited to uncertainty propagation. One drawback is that mixtures of Gaussian random variables are generally not Gaussian, so that the usual interpretations and conditioning on Gaussian processes will not hold any more. The so-called *Gaussian Process Regression* (Williams and Rasmussen [1996], Rasmussen and Williams [2006]) will have to be adapted to this new setting.

The paper structure is as follows: in Section 2, we present the formalization of the considered model and how we can model data given on areas, uncertainty on this data, multidimensional outputs, and missing outputs. In Section 3, we detail some prediction results to predict several outputs on specific areas. At last, in Section 4 we give some numerical illustrations to show the usefulness and the performance of the method.

2 Spatial model

2.1 Inputs

We model first the input locations in dimension d , in order to define the studied territory and some of its subsets.

Definition 1 (Territory). *Let d be a positive integer corresponding to a dimension. A territory and grains inside this territory are defined as follows:*

- A **territory** is a subset χ of \mathbb{R}^d .
- A **grain** is any nonempty subset $g \subseteq \chi$.
- A **point** is any element $x \in \chi$.

As an example, if points are modelled by a couple (*latitude, longitude*), then χ can be the set of all (*latitude, longitude*) corresponding to a country, in that case $d = 2$ and $\chi \subset \mathbb{R}^2$. A grain may correspond, for example, to a specific city, to a specific land plot, or to a specific building. Previous Kriging models refer to blocks or areas for sets of points that are disjoint and those authors are not interested in the family itself (see for instance Kyriakidis [2004]). We develop in Appendix A some considerations about those families that arise when relaxing the disjunction constraint.

In order to deal with several sets of grains, we also define a granularity. This is not compulsory to understand the model, but it may ease further applications. For instance, in geographic information, granularities may be the set of land plots, the set of cities, the set of buildings, etc.

In many practical situations, considered grains may have non empty intersections. They may also come from different datasets, at different scales. We have chosen below some definitions that help the manipulation of these sets of grains, especially in problematic cases with non-empty intersections and multiscale data. The construction of set of grains with empty intersections will follow in a natural way.

Definition 2 (Granularities). *In a given territory χ , a **granularity** $\mathcal{G} = \{g_1, g_2, \dots\}$ is a finite set of grains, not necessarily pairwise distinct.*

In Appendix A, we define several tools useful to work with granularities: non-overlapping granularity, granularity order, insertion operator, maximal non-overlapping granularity.

2.2 Outputs

We now aim at defining variables that can be observed at a specific point $x \in \chi$ as well as at the scale of a grain.

At each input location $x \in \chi$, the outputs $\mathbf{Y}(x)$ may represent, say, the insulation level of a building, the households income, the square meter price, the building'energy efficiency... We consider here $p \in \mathbb{N}^*$ output variables. At a larger scale, it is difficult to specify what should represent the outputs over a whole space, e.g. over a city. We will detail it in this section.

Definition 3 (Outputs). *Let \mathcal{G} be a granularity. The outputs are defined over points and grains of \mathcal{G} as follows:*

- \mathbf{Y} is a p -dimensional multivariate random field over χ denoted:

$$\forall x \in \chi, \mathbf{Y}(x) := (Y_1(x), \dots, Y_p(x))^{\top} \in \mathbb{R}^p$$

- For each $g \in \mathcal{G}$, a p -dimensional real random vector $\mathbf{Y}(g)$ is defined to be the value of \mathbf{Y} at a random location $X_g \in g$:

$$\forall g \in \mathcal{G}, \mathbf{Y}(g) := \mathbf{Y}(X_g) \in \mathbb{R}^p$$

For a given granularity \mathcal{G} , we assume that the set of random variables $\{X_g : g \in \mathcal{G}\}$, is defined and known, and that the dependence structure between those random variables is also known. We assume furthermore that these random variables are independent from the random field \mathbf{Y} .

The above definition gives a specific answer to the problem (P1) in the introduction. Other models can be found in the literature. It is common to use averaged variables to describe outputs over a larger scale, e.g. the average household income over a city $\bar{\mathbf{Y}}(g) := \int_g \mathbf{Y}(x) dF_g(x)$ with a suitable cumulative distribution function F_g , possibly discrete (see for instance Equation (1) in Gotway and Young [2002] or in Kyriakidis [2004] for equally weighted linear combination, i.e. uniformly distributed F_g). However, such an approach has some drawbacks. Firstly, it reduces the dispersion of the variable as the scale grows. Secondly, the application of any highly convex function h would induce a large bias, as $\mathbb{E}[h(\mathbf{Y}(g))] \neq h(\mathbb{E}[\mathbf{Y}(g)])$. The average value is mostly interesting as an unbiased estimator of the expectation, precisely because its dispersion is small but it strongly shrinks the dispersion which has some adverse effects. Indeed, underestimating the dispersion of an output random variable is an adverse effect when we plan to feed a machine learning algorithm with this data. Such an algorithm would be jeopardized by an averaged variable resulting in predicting a unique value instead of explaining the variance of the dataset. Let us take life-expectancy for instance, the averaged value might be very similar from one city to another but with very different dispersions, hiding social inequalities. In addition, convexity bias is a problem when we transform variables or build composite indicators. We can illustrate those problems as follows:

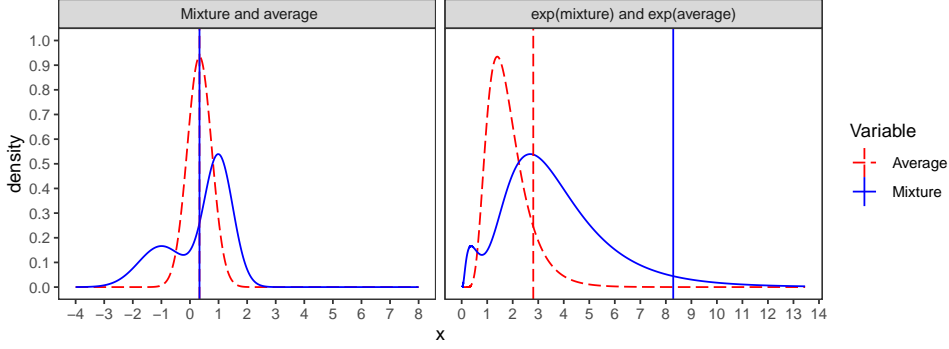


Figure 1: *Comparison of mixture and average.* **Left:** Density of a mixture Y_{mixture} picking Y_a with probability $\frac{1}{3}$ and picking Y_b with probability $\frac{2}{3}$ (blue plain line), and density of the average $Y_{\text{average}} = \frac{1}{3}Y_a + \frac{2}{3}Y_b$ (red dashed line), where $Y_a \sim N(-1, 0.8)$ and $Y_b \sim N(1, 0.5)$. The common mean of both distributions appears as a vertical line. The mixture is no longer Gaussian and exhibits a higher dispersion. **Right:** Density of $\exp(Y_{\text{mixture}})$ (blue plain line), and density of $\exp(Y_{\text{average}})$ (red dashed line). The two different means appear as vertical lines. The dispersion of the exponential of the mixture is higher than the one of the average, and the difference between both means illustrates the convexity bias

- **Effects on dispersion** Let us assume that we have a grain $g = \{x_i : i \in \llbracket 1, n \rrbracket\}$, $p = 1$ and consider mutually independent r.v. $\forall i, \mathbf{Y}(x_i) \sim \mathcal{N}(\mu, \sigma^2)$ for given μ and σ . The average of those variables (usually called sample mean) is $\bar{\mathbf{Y}}(g) := \frac{1}{n} \sum_{i=1}^n \mathbf{Y}(x_i) \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$. Considering a discrete uniform distribution for X_g , the mixture variable is following the same law as any point of the grain: $\mathbf{Y}(X_g) \sim \mathcal{N}(\mu, \sigma^2)$. This is precisely what one would expect when assessing for instance the salaries distribution over a territory knowing the distribution at a smaller scale.
- **Convexity bias** Consider now i.i.d. random variables $\forall i \in \llbracket 1, n \rrbracket, \mathbf{Y}(x_i) \sim \mathcal{U}(a, b)$ (uniform distribution on $[a, b]$) for given $a, b \geq 0$. We define $h(x) := x^2$ and $\bar{\mathbf{Y}}(g) := \frac{1}{n} \sum_{i=1}^n \mathbf{Y}(x_i)$. When n grows, $h(\bar{\mathbf{Y}}(g))$ depends upon n and quickly converges towards $h(\frac{a+b}{2}) = \frac{a^2+2ab+b^2}{4}$. However, $\mathbf{Y}(X_g) \sim \mathcal{U}(a, b)$ does not depend on n and $\mathbb{E}[h(\mathbf{Y}(X_g))] = \frac{a^2+ab+b^2}{3}$, resulting in a bias:

$$\mathbb{E}[h(\mathbf{Y}(X_g))] - \mathbb{E}[h(\bar{\mathbf{Y}}(g))] \xrightarrow{n \rightarrow +\infty} \frac{(b-a)^2}{12} = \mathbb{E}[h(\mathbf{Y}(x_1))] - h(\mathbb{E}[\mathbf{Y}(x_1)])$$

In this case, taking the averaged variable $\bar{\mathbf{Y}}(g)$ instead of $\mathbf{Y}(X_g)$ would result not only in losing most of the dispersion of $h(\mathbf{Y}(X_g))$ but also in underestimating its mean, as $\mathbb{E}[h(\mathbf{Y}(X_g))] > h(\mathbb{E}[\bar{\mathbf{Y}}(g)])$.

We compare mean and mixture distributions of two Gaussian random variables in Figure 1. Mean and mixture have same expectation $1/3$ but their dispersion is different (Figure 1, left). The application of a convex function accentuates this dispersion difference, but also creates a differences between the expectations, which is called here the convexity bias (Figure 1, right).

The originality of the present work is to use mixtures to define $\{\mathbf{Y}(g) : g \in \mathcal{G}\}$. At the scale of a grain g , $\mathbf{Y}(g)$ is defined to be equal to $\mathbf{Y}(X_g)$ the value of \mathbf{Y} at a random location $X_g \in g$. If the joint distribution of $\{X_g \in \chi : g \in \mathcal{G}\}$ is known, then the joint distribution of $\mathbf{Y}(g)$ will be deduced. Or at least, if the moments of order one and cross moments of order two of $\{X_g \in \chi : g \in \mathcal{G}\}$ exist and are known, then one would be able to determine expectation and cross covariances of $\{\mathbf{Y}(g) : g \in \mathcal{G}\}$. In the rest of the paper, we assume that first two moments of $\{X_g \in \chi : g \in \mathcal{G}\}$ and $\{\mathbf{Y}(g) : g \in \mathcal{G}\}$ exist.

In the following proposition, we show that if means and covariances of $\mathbf{Y}(x)$ are known, then mean and covariances of $\mathbf{Y}(g)$ can be computed. This will help in particular addressing the uncertainty propagation problem (P2) in the introduction.

Proposition 1 (Mean and covariances of $\mathbf{Y}(g)$). *From Definition 3, we derive the following results:*

(i) *For any grain $g \in \mathcal{G}$ and any index $i \in \llbracket 1, p \rrbracket$, assuming that for all $x \in g$ we know $\mu_i(x) := \mathbb{E}[Y_i(x)]$, we have:*

$$\mu_i(g) := \mathbb{E}[Y_i(g)] = \mathbb{E}[\mu_i(X_g)] \quad (2)$$

(ii) *For any two grains g, g' in \mathcal{G} and any two indices $i, j \in \llbracket 1, p \rrbracket$, assuming that for all $x \in g, x' \in g'$ we know $k_{i,j}(x, x') := \text{Cov}[Y_i(x), Y_j(x')]$, we have:*

$$k_{i,j}(g, g') := \text{Cov}[Y_i(g), Y_j(g')] = \mathbb{E}[k_{i,j}(X_g, X_{g'})] + \text{Cov}[\mu_i(X_g), \mu_j(X_{g'})] \quad (3)$$

In particular, $k_{i,i}(g, g) = \text{Cov}[Y_i(g), Y_i(g)] = \mathbb{V}[Y_i(g)] = \mathbb{E}[k_{i,i}(X_g, X_g)] + \mathbb{V}[\mu_i(X_g)]$.

Proof. (i) is a direct application of the conditional expectation formula where $Y_i(g)$ is the result of conditioning $Y_i(x)$ with X_g . (ii) derives from the conditional covariance (variance) formula, after conditioning by the joint random vector $(X_g, X_{g'})$ (random variable X_g). \square

For example, $\{X_g \in \chi : g \in \mathcal{G}\}$ can be mutually independent random variables. The distribution of each X_g can be assumed to be uniform over a subset of g , whether it is a continuous or discrete distribution, so that all quantities in Equations (2) and (3) can be computed easily, for all $g, g' \subset \mathcal{G}$.

Note that $\text{Cov}[\mu_i(X_g), \mu_j(X_{g'})] = 0$ in the case where $\mu_i(x)$ is constant over any one of the grains g or g' or in the case where X_g and $X_{g'}$ are independent. Also note that this framework yields the expected result that if a grain is restricted to a point, then the output of this grain is the same as the output of the underlying point.

Remark 1 (Comparison with average – block-to-block covariances). *Define $\bar{Y}_i(g) = \mathbb{E}[Y_i(X_g) | \{Y_i(x), x \in g\}] = \int_g Y_i(x) dF_g(x)$, with F_g the cdf of the (possibly discrete) random variable X_g , $i \in \llbracket 1, p \rrbracket$. One can check that with this setting the mean of the mixture $Y_i(g)$ and the average $\bar{Y}_i(g)$ are identical:*

$$\mathbb{E}[Y_i(g)] = \bar{Y}_i(g).$$

For the covariances, when X_g and $X_{g'}$ are two independent random variables, one can check that

$$\mathbb{E}[k_{i,j}(X_g, X_{g'})] = \text{Cov}[\bar{Y}_i(g), \bar{Y}_j(g')]$$

However

$$\mathbb{E}[k_{i,j}(X_g, X_g)] \neq \text{Cov}[\bar{Y}_i(g), \bar{Y}_j(g)]$$

because the independence assumption does not hold any more. Finally, $\mathbb{V}[Y_i(g)] \neq \mathbb{V}[\bar{Y}_i(g)]$, even in the specific case where $\forall i, j, g, g', \text{Cov}[\mu_i(X_g), \mu_j(X_{g'})] = 0$. One retrieves here the difference between a mixture and an average, where the mixture can exhibit a higher dispersion.

Note that it has been proved that reducing a grain to its centroid is distorting the covariances as compared to taking into account its shape and size, even if grains are of same shape and size. See for instance Goovaerts [2008] for a discussion on this issue through distances between grains. Distortion is increased for short distances which are precisely the interesting ones for covariances. Using centroids can also become a non-sense since centroids may very well lie outside the grain itself for non-convex grains. The cost of ruling out such an approximation is that in certain cases, we will have to estimate the point-support covariance from only aggregated observations.

Remark 2 (Comparison with average – aggregation effect). *Let g and g' be two non-overlapping grains (i.e. $g \cap g' = \emptyset$). From the two distributions of X_g and $X_{g'}$, one can define a mixture r.v. $X_{g \cup g'}$. From $X_{g \cup g'}$, one can thus build $\mathbf{Y}(X_{g \cup g'})$ as a mixture of $\mathbf{Y}(X_g)$ and $\mathbf{Y}(X_{g'})$. From $\bar{\mathbf{Y}}(g)$ and $\bar{\mathbf{Y}}(g')$, one can also build $\bar{\mathbf{Y}}(g \cup g')$, but the latter aggregation, which uses average, creates many problems. As presented in the introduction, this has been discussed in Gotway and Young [2002]. The modifiable areal unit problem (MAUP) is focused on the aggregation effect and it turns out that correlation between output variables at grain level is modified in a controllable way when aggregating grains (aggregation effect) due to “the smoothing effect that results from averaging” Gotway and Young [2002]. A quantification method for the MAUP has been recently proposed by Briz-Redón [2022].*

In the Example 1 below, one investigates the impact of overlapping granularities. In many cases, the overlaps impact is limited. In situations where this impact can be important, one can use the construction of non-overlapping granularity presented in Proposition 5 (see Appendix A).

Example 1 (Overlapping granularity). Consider two overlapping grains g and g' , with nonempty intersection $g_0 = g \cap g'$. We want to compare the situation where X_g is dependent on $X_{g'}$ with a situation of independence.

- *Case of dependence.* We define random locations X_{g_0} , $X_{g \setminus g_0}$, $X_{g' \setminus g_0}$ and two Bernoulli random variables B and B' . We assume that those five random variables are mutually independent. Let:

$$\begin{cases} X_g &= BX_{g_0} + (1-B)X_{g \setminus g_0} \\ X_{g'} &= B'X_{g_0} + (1-B')X_{g' \setminus g_0} \end{cases} \quad (4)$$

- *Case of independence.* We introduce here $X_{g_0}^\perp$ an independent copy of X_{g_0} , independent from X_{g_0} , $X_{g \setminus g_0}$, $X_{g' \setminus g_0}$, B and B' . Let:

$$\begin{cases} X_g &= BX_{g_0} + (1-B)X_{g \setminus g_0} \\ X_{g'}^\perp &= B'X_{g_0}^\perp + (1-B')X_{g' \setminus g_0} \end{cases} \quad (5)$$

Let Δ be the covariance difference due to the dependence structure of X_g and $X_{g'}$,

$$\Delta := \text{Cov} [Y_i(X_g), Y_j(X_{g'}^\perp)] - \text{Cov} [Y_i(X_g), Y_j(X_{g'})]. \quad (6)$$

Then setting $\rho_{\max} = \sup \{|k_{i,j}(x, x) - k_{i,j}(x, x')| : x \in g_0, x' \in g_0\}$, assuming that

$$\forall x \in g \cup g', \quad \begin{cases} \mu_i(x) = \mu_i(g) = \mu_i(g') \\ \mu_j(x) = \mu_j(g) = \mu_j(g') \end{cases}$$

one can show that:

$$|\Delta| \leq \mathbb{P}[B = B' = 1] \mathbb{P}[X_{g_0} \neq X_{g_0}^\perp] \rho_{\max}. \quad (7)$$

The variation due to the common dependence structure on the overlap can be significant if all of the three factors are not negligible. This shows in particular that overlapping grains are not too problematic, when means are identical, if the probability of selecting the intersection g_0 for both grain is small, or if the probability of selecting different points in the intersection is small.

Proof of the results in Example 1. Under given assumptions on the means μ_i and μ_j , Applying the total covariance formula on $\text{Cov} [Y_i(X_g), Y_j(X_{g'}^\perp)]$ and $\text{Cov} [Y_i(X_g), Y_j(X_{g'})]$, we get

$$\Delta = \mathbb{E} [\text{Cov} [Y_i(X_g), Y_j(X_{g'}^\perp)] | (B, B')] - \mathbb{E} [\text{Cov} [Y_i(X_g), Y_j(X_{g'})] | (B, B')],$$

and the difference is non zero in the only case where $B = B' = 1$, so that using independence,

$$\Delta = \mathbb{P}[B = B' = 1] (\mathbb{E} [\text{Cov} [Y_i(X_{g_0}), Y_j(X_{g_0})]) - \mathbb{E} [\text{Cov} [Y_i(X_{g_0}), Y_j(X_{g_0}^\perp)])]$$

The parenthesis vanishes in any conditional cases where $X_{g_0}^\perp = X_{g_0}$, and in other cases, the conditional difference is bounded by ρ_{\max} , hence the result. \square

3 Prediction

To define our **learning data** we now assume that the output is partially known on a set of grains:

For $(i_1, \dots, i_n) \in \llbracket 1, p \rrbracket^n$ and $g_1, \dots, g_n \in \mathcal{G}$ we know n random variables:

$$\underline{\mathbf{Y}} = (Y^1, \dots, Y^n)^\top \quad \text{with} \quad Y^j = Y_{i_j}(g_j) \quad \text{for} \quad j \in \llbracket 1, n \rrbracket$$

As an example, if k observations of the whole random vector $\mathbf{Y}(g_j)$ are conducted for $j \in \llbracket 1, k \rrbracket$, then setting $n = k \cdot p$ allows storing the whole available information:

$$\underline{\mathbf{Y}} = (Y_1(X_{g_1}), \dots, Y_p(X_{g_1}), \dots, Y_1(X_{g_j}), \dots, Y_p(X_{g_j}), \dots, Y_1(X_{g_k}), \dots, Y_p(X_{g_k}))^\top. \quad (8)$$

If some observations are incomplete, that is to say some components of \mathbf{Y}_{g_j} are missing for some j , then $\underline{\mathbf{Y}}$ will be a subvector of \mathbf{Y} as given in Equation (8). The problem (P3) in the introduction can thus be treated easily with this formalism.

We want to predict the output $\mathbf{Y}(g)$ for a new grain $g \subset \chi$. Note that in the specific case where this new grain is reduced to a single location $g = \{x\}$, this prediction corresponds to the one of $\mathbf{Y}(x)$. As $\mathbf{Y}(g) = (Y_1(g), \dots, Y_p(g))^\top$, the problem amounts to predicting $Y_i(g)$ for all $i \in \llbracket 1, p \rrbracket$.

For some $g \subset \chi$ and some $i \in \llbracket 1, p \rrbracket$, we want to **predict** $Y_i(g)$ from a **learning set** of observations $\underline{\mathbf{Y}}$.

We denote:

$$\begin{aligned} \underline{\boldsymbol{\mu}} &:= \mathbb{E}[\underline{\mathbf{Y}}] && \in \mathbb{R}^n \\ \mathbf{K} &:= \left(\text{Cov} \left[Y^j, Y^{j'} \right] \right)_{j, j' \in \llbracket 1, n \rrbracket} && \in \mathcal{S}_n^+(\mathbb{R}) \text{ semi-definite positive } n \times n \text{ matrix} \\ \mathbf{h}_i(g) &:= \left(\text{Cov} \left[Y^j, Y_i(g) \right] \right)_{j \in \llbracket 1, n \rrbracket} && \in \mathbb{R}^n \end{aligned}$$

In the following, we assume that \mathbf{K} is invertible. Note that if the expectations of $Y_i(x)$ and covariances between $Y_i(x)$ and $Y_j(x')$ are known for all $i, j \in \llbracket 1, p \rrbracket$, $x, x' \in \chi$, as is the case in usual Kriging assumptions, $\underline{\boldsymbol{\mu}}$, \mathbf{K} and $\mathbf{h}_i(g)$ can be computed using Proposition 1.

We look for the optimal weights $\boldsymbol{\alpha}_i(g) = (\alpha_i^1(g), \dots, \alpha_i^n(g)) \in \mathbb{R}^n$ associated with the best linear unbiased predictor $M_i(g)$ of $Y_i(g)$ such that:

$$M_i(g) = \sum_{j=1}^n \alpha_i^j(g) Y^j = \boldsymbol{\alpha}_i(g)^\top \underline{\mathbf{Y}}. \quad (9)$$

where the model is optimized by minimizing a quadratic error over all unbiased linear predictors:

$$\boldsymbol{\alpha}_i(g) \in \arg \min_{\boldsymbol{\alpha} \in \mathbb{R}^n} \mathbb{E} \left[(Y_i(g) - \boldsymbol{\alpha}^\top \underline{\mathbf{Y}})^2 \right] \quad (10)$$

Given the optimal predictor $M_i(g)$, the resulting errors are denoted:

$$\begin{cases} \epsilon_i(g) & := Y_i(g) - M_i(g) \\ c_{i,j}(g, g') & := \mathbb{E}[\epsilon_i(g) \epsilon_j(g')] \\ v_i(g) & := c_{i,i}(g, g) \end{cases} \quad (11)$$

Remark 3 (Spaces of linear unbiased predictors). *It is interesting for the understanding of the problem to give it a geometrical approach. Let us denote $F_i(g)$ the set of linear unbiased predictors of $Y_i(g)$ given an observation vector $\underline{\mathbf{Y}}$. With previous notations, it means that:*

$$F_i(g) = \{ \boldsymbol{\alpha}^\top \underline{\mathbf{Y}} : \mu_i(g) = \boldsymbol{\alpha}^\top \underline{\boldsymbol{\mu}} \}$$

And similarly, we denote:

$$\begin{aligned} F &:= \{ \boldsymbol{\alpha}^\top \underline{\mathbf{Y}} : \boldsymbol{\alpha} \in \mathbb{R}^n \} \\ F_0 &:= \{ \boldsymbol{\alpha}^\top \underline{\mathbf{Y}} : \boldsymbol{\alpha}^\top \underline{\boldsymbol{\mu}} = 0 \} \end{aligned}$$

One can note that F_0 is a subspace of F of dimension $\dim(F) - 1$. Moreover $F_0 + F_i(g) = F_i(g)$, meaning that $F_i(g)$ is an affine subspace of F having F_0 for underlying vector space. But it also means that some spaces are parallel:

$$\forall i, j \in \llbracket 1, p \rrbracket, \forall g, g' \in \chi, F_i(g) \parallel F_j(g')$$

Now, given that we are minimizing the quadratic error between $Y_i(g)$ and $M_i(g)$ which can be seen as a distance, the optimization process is geometrically a projection of $Y_i(g)$ on $F_i(g)$. This approach is illustrated in Figure 2.

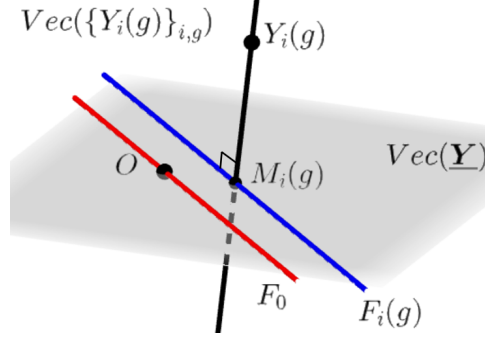


Figure 2: Geometrical interpretation of the prediction process.

The following proposition gives an optimal predictor that can be computed under the minimal assumptions of Proposition 1: given the first two moments of random variables $\{X_g : g \in \mathcal{G}\}$, all components of $\underline{\mu}$, \mathbf{K} and $\mathbf{h}_i(x)$ can be computed.

Proposition 2 (Mixture Kriging prediction). *Given a set of observations $\underline{\mathbf{Y}}$, for any $g \in \chi$, and in particular for a single point $g = \{x\}$, for any $i \in \llbracket 1, p \rrbracket$, the weights $\alpha_i(g)$ yielding the best linear unbiased predictor (BLUP) of $Y_i(g)$ and the associated cross errors are as follows:*

(i) **Simple Mixture Kriging.** If $\underline{\mu} = (0, \dots, 0)^\top$ and $\mu_i(g) = 0$ then

$$\begin{cases} \alpha_i(g) &= \mathbf{K}^{-1} \mathbf{h}_i(g) \\ c_{i,j}(g, g') &= k_{i,j}(g, g') - \mathbf{h}_i(g)^\top \mathbf{K}^{-1} \mathbf{h}_j(g') \end{cases}$$

(ii) **Ordinary mixture Kriging.** If $\underline{\mu} \neq (0, \dots, 0)^\top$ then the condition for unbiasedness writes $\mu_i(g) = \alpha_i(g)^\top \underline{\mu}$ and

$$\begin{cases} \alpha_i(g) &= \mathbf{K}^{-1} (\mathbf{h}_i(g) + \lambda_i(g) \underline{\mu}) \quad \text{where } \lambda_i(g) = \frac{\mu_i(g) - \underline{\mu}^\top \mathbf{K}^{-1} \mathbf{h}_i(g)}{\underline{\mu}^\top \mathbf{K}^{-1} \underline{\mu}} \\ c_{i,j}(g, g') &= k_{i,j}(g, g') - \mathbf{h}_i(g)^\top \mathbf{K}^{-1} \mathbf{h}_j(g') + \lambda_i(g) \lambda_j(g) \underline{\mu}^\top \mathbf{K}^{-1} \underline{\mu} \end{cases}$$

Proof. For given $i \in \llbracket 1, p \rrbracket$ and $g \subseteq \chi$, let $M_\alpha = \alpha^\top \underline{\mathbf{Y}}$ be a linear predictor of $Y_i(g)$, where $\alpha = (\alpha^1, \dots, \alpha^n)$ is a vector of weights, and denote the associated error $v_i(g, \alpha) := \mathbb{E}[(Y_i(g) - M_\alpha)^2]$, then:

$$\begin{aligned} v_i(g, \alpha) &= \mathbb{E} \left[(\alpha^\top \underline{\mathbf{Y}} - Y_i(g))^2 \right] \\ &= \mathbb{E} \left[\alpha^\top \underline{\mathbf{Y}} \underline{\mathbf{Y}}^\top \alpha - 2Y_i(g) \alpha^\top \underline{\mathbf{Y}} + Y_i(g)^2 \right] \\ &= \alpha^\top \mathbf{K} \alpha + \alpha^\top \underline{\mu} \underline{\mu}^\top \alpha - 2\alpha^\top (\mathbf{h}_i(g) + \underline{\mu} \mu_i(g)) + \mathbb{V}[Y_i(g)] + \mu_i(g)^2. \end{aligned}$$

(i) If $\underline{\mu} = (0, \dots, 0)^\top$ and $\mu_i(g) = 0$ then

$$v_i(g, \alpha) = \alpha^\top \mathbf{K} \alpha - 2\alpha^\top \mathbf{h}_i(g) + \mathbb{V}[Y_i(g)].$$

By differentiation over each component of α ,

$$\frac{\partial v_i(g, \alpha)}{\partial \alpha} := \left(\frac{\partial v_i(g, \alpha)}{\partial \alpha^j} \right)_{j \in \llbracket 1, p \rrbracket} = 2\mathbf{K} \alpha - 2\mathbf{h}_i(g).$$

Without constraints, this value should be null at any extremum, and thus the optimal vector of weights is

$$\alpha_i(g) = \mathbf{K}^{-1} \mathbf{h}_i(g).$$

Since \mathbf{K} is symmetric positive, this only extremum is a minimum.

(ii) If $\underline{\boldsymbol{\mu}} \neq (0, \dots, 0)^\top$ then the condition for unbiasedness writes $\mu_i(g) = \boldsymbol{\alpha}^\top \underline{\boldsymbol{\mu}}$ by linearity of expectation.

$v_i(g, \boldsymbol{\alpha})$ rewrites again:

$$v_i(g, \boldsymbol{\alpha}) = \boldsymbol{\alpha}^\top \mathbf{K} \boldsymbol{\alpha} - 2\boldsymbol{\alpha}^\top \mathbf{h}_i(g) + \mathbb{V}[Y_i(g)].$$

We introduce the Lagrangian operator:

$$\mathcal{L}(\boldsymbol{\alpha}, \lambda) = v_i(g, \boldsymbol{\alpha}) - 2\lambda(\boldsymbol{\alpha}^\top \underline{\boldsymbol{\mu}} - \mu_i(g)).$$

We are minimizing a quadratic function over a single affine equality constraint. A necessary optimality condition is:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\alpha}}(\boldsymbol{\alpha}, \lambda) = 0,$$

that is to say:

$$2\mathbf{K}\boldsymbol{\alpha} - 2\mathbf{h}_i(g) - 2\lambda\underline{\boldsymbol{\mu}} = 0,$$

and therefore the optimal weights are

$$\boldsymbol{\alpha}_i(g) = \mathbf{K}^{-1}(\mathbf{h}_i(g) + \lambda\underline{\boldsymbol{\mu}}).$$

The unbiasedness condition is:

$$\underline{\boldsymbol{\mu}}^\top (\mathbf{K}^{-1}(\mathbf{h}_i(g) + \lambda\underline{\boldsymbol{\mu}})) = \mu_i(g),$$

so that

$$\lambda_i(g) = \frac{\mu_i(g) - \underline{\boldsymbol{\mu}}^\top \mathbf{K}^{-1} \mathbf{h}_i(g)}{\underline{\boldsymbol{\mu}}^\top \mathbf{K}^{-1} \underline{\boldsymbol{\mu}}}.$$

Therefore this only solution is a minimum of $v_i(g, \boldsymbol{\alpha})$.

Let us consider now the cross-errors:

$$c_{i,j}(g, g') = \mathbb{E}[(Y_i(g) - M_i(g))(Y_j(g') - M_j(g'))].$$

Due to unbiasedness condition, it means that:

$$\begin{aligned} c_{i,j}(g, g') &= \text{Cov}[Y_i(g) - M_i(g), Y_j(g') - M_j(g')] \\ &= \text{Cov}[Y_i(g), Y_j(g')] - \text{Cov}[Y_i(g), M_j(g')] - \text{Cov}[M_i(g), Y_j(g')] + \text{Cov}[M_i(g), M_j(g')] \\ &= \text{Cov}[Y_i(g), Y_j(g')] - \text{Cov}[Y_i(g), \boldsymbol{\alpha}_j(g')^\top \underline{\mathbf{Y}}] - \text{Cov}[\boldsymbol{\alpha}_i(g)^\top \underline{\mathbf{Y}}, Y_j(g')] + \text{Cov}[\boldsymbol{\alpha}_i(g)^\top \underline{\mathbf{Y}}, \boldsymbol{\alpha}_j(g')^\top \underline{\mathbf{Y}}]. \end{aligned}$$

Which rewrites:

$$c_{i,j}(g, g') = k_{i,j}(g, g') - \boldsymbol{\alpha}_j(g')^\top \mathbf{h}_i(g) - \boldsymbol{\alpha}_i(g)^\top \mathbf{h}_j(g') + \boldsymbol{\alpha}_i(g)^\top \mathbf{K} \boldsymbol{\alpha}_j(g'). \quad (12)$$

Note that equation (12) is true for any linear unbiased predictor.

Which, in the case of simple mixture Kriging, simplifies into:

$$c_{i,j}(g, g') = k_{i,j}(g, g') - \mathbf{h}_i(g)^\top \mathbf{K}^{-1} \mathbf{h}_j(g').$$

And in the case of ordinary mixture Kriging:

$$c_{i,j}(g, g') = k_{i,j}(g, g') - \mathbf{h}_i(g)^\top \mathbf{K}^{-1} \mathbf{h}_j(g') + \lambda_i(g) \lambda_j(g) \underline{\boldsymbol{\mu}}^\top \mathbf{K}^{-1} \underline{\boldsymbol{\mu}}.$$

The expressions of $v_i(g) = c_{i,i}(g, g)$ in both cases follow immediately. \square

Remark 4 (Ordinary Mixture Kriging and mean estimation). *Regarding ordinary mixture Kriging, assuming that all random variables $Y_i(g)$ have the same unknown expectation, setting $\mathbf{1}_n = (1, \dots, 1)^\top$, we have:*

$$\boldsymbol{\alpha}_i(g) = \mathbf{K}^{-1} \left(\mathbf{h}_i(g) + \frac{1 - \mathbf{1}_n^\top \mathbf{K}^{-1} \mathbf{h}_i(g)}{\mathbf{1}_n^\top \mathbf{K}^{-1} \mathbf{1}_n} \mathbf{1}_n \right),$$

and setting

$$\hat{m}(g) := \frac{\mathbf{1}_n^\top \mathbf{K}^{-1} \mathbf{Y}}{\mathbf{1}_n^\top \mathbf{K}^{-1} \mathbf{1}_n},$$

we can write:

$$M_i(g) = \hat{m}(g) + \mathbf{h}_i(g)^\top \mathbf{K}^{-1} (\mathbf{Y} - \mathbf{1}_n \hat{m}(g)),$$

therefore $\hat{m}(g)$ is an unbiased estimator of the unknown common expectation. \hat{m} can be compared with usual sample mean for independent observations $\bar{\mathbf{Y}} = \frac{\mathbf{1}_n^\top \mathbf{Y}}{\mathbf{1}_n^\top \mathbf{1}_n}$.

Proposition 3 (Noisy observations). *Let us consider the case where for a given $x \in \chi$ we can only observe $\tilde{Y}_i(x) = Y_i(x) + \epsilon_i(x)$ where $\epsilon_i(x)$ is independent from any $Y_j(x')$. We denote the resulting noisy outputs, observations and covariances:*

$$\begin{aligned} \tilde{Y}_i(g) &:= \tilde{Y}_i(X_g) = Y_i(g) + \epsilon_i(g) \\ \tilde{Y}^j &:= \tilde{Y}_{i_j}(X_{g_j}) = Y^j + \epsilon^j \\ \eta_{i,j}(x, x') &:= \text{Cov}[\epsilon_i(x), \epsilon_j(x')] \end{aligned}$$

Then covariance between 2 grains outputs is:

$$\tilde{k}_{i,j}(g, g') := \text{Cov}[\tilde{Y}_i(g), \tilde{Y}_j(g')] = k_{i,j}(g, g') + \mathbb{E}[\eta_{i,j}(X_g, X_{g'})]$$

Therefore observations covariance matrix writes:

$$\begin{aligned} \tilde{\mathbf{K}} &:= \left(\text{Cov}[\tilde{Y}^j, \tilde{Y}^{j'}] \right)_{j, j' \in \llbracket 1, n \rrbracket} \\ \tilde{\mathbf{K}} &= \mathbf{K} + \left(\text{Cov}[\epsilon^j, \epsilon^{j'}] \right)_{j, j' \in \llbracket 1, n \rrbracket} \\ \tilde{\mathbf{K}} &= \mathbf{K} + \mathbf{K}_\epsilon \end{aligned}$$

And covariance vector between observations and a new grain writes:

$$\begin{aligned} \tilde{\mathbf{h}}_i(g) &:= \left(\text{Cov}[Y^j + \epsilon^j, Y_i(g) + \epsilon_i(g)] \right)_{j \in \llbracket 1, n \rrbracket} \\ \tilde{\mathbf{h}}_i(g) &= \mathbf{h}_i(g) + \left(\mathbb{E}[\eta_{i,j,i}(X_{g_j}, X_g)] \right)_{j \in \llbracket 1, n \rrbracket} \\ \tilde{\mathbf{h}}_i(g) &= \mathbf{h}_i(g) + \mathbf{h}_{\epsilon,i}(g) \end{aligned}$$

Typically, we can assume that $\mathbb{E}[\eta_{i,j}(X_g, X_{g'})] = \mathbb{1}_{\{i=j\}} \mathbb{1}_{\{g=g'\}} \eta_{i,i}(g, g)$. In which case \mathbf{K}_ϵ is a diagonal matrix and $\mathbf{h}_{\epsilon,i}(g)$ is null as long as g is not among the observed grains.

Remark 5 (Gaussian Singleton case). *Assume that $\{\mathbf{Y}(x) : x \in \chi\}$ is a vector-valued Gaussian random field and that each X_g is Dirac distributed. This last condition holds in particular when each grain is restricted to one singleton point. In this Gaussian case, one retrieves the usual Simple Kriging and Ordinary Kriging predictors, as defined for example in Rasmussen and Williams [2006]. This remark also holds for the next Proposition 4. In this sense, the Mixture Kriging results presented here can be seen as a generalization of the usual Kriging interpolation.*

Proposition 4 (Cross-errors and conditional covariances). *Consider the assumption*

$$(A) : \quad \forall i \in \llbracket 1, p \rrbracket, \quad \forall g \in \mathcal{G}, \quad M_i(g) = \mathbb{E}[Y_i(g) | \mathbf{Y}].$$

This is for example the case when $\{\mathbf{Y}(x) : x \in \chi\}$ is a vector-valued Gaussian random field and when each X_g is Dirac distributed (see Remark 5). In this setting, under assumption (A), one can show that cross errors for both Simple Mixture Kriging and Ordinary Mixture Kriging are

$$c_{i,j}(g, g') = \mathbb{E}[\text{Cov}[Y_i(g), Y_j(g') | \mathbf{Y}]]. \quad (13)$$

If $\text{Cov}[Y_i(g), Y_j(g') | \mathbf{Y}]$ does not depend on \mathbf{Y} , as it is the case for conditional Gaussian vectors, Equation (13) simplifies: $\mathbb{E}[\text{Cov}[Y_i(g), Y_j(g') | \mathbf{Y}]] = \text{Cov}[Y_i(g), Y_j(g') | \mathbf{Y}]$.

Proof. The proof uses a classical approach on orthogonality of Best Linear Unbiased Predictors. It is presented here in three steps. The proof can be simplified in the Simple Mixture Kriging setting.

- First, given the notations introduced in Remark 3, let $\delta \in F_0$ be a non-zero vector and β a real number.

Let $M_i^\beta(g) := M_i(g) + \beta \delta \in F_i(g)$. Recall that $\epsilon_i(g) := Y_i(g) - M_i(g)$ and $v_i(g) := \mathbb{E}[(\epsilon_i(g))^2]$.

We have:

$$\mathbb{E}[(Y_i(g) - M_i^\beta(g))^2] = v_i(g) - 2\beta \mathbb{E}[\epsilon_i(g) \delta] + \beta^2 \mathbb{E}[\delta^2].$$

The minimum value of this polynomial expression is reached for:

$$\beta_0 = \frac{\mathbb{E}[\epsilon_i(g) \delta]}{\mathbb{E}[\delta^2]}.$$

Since the only optimal point is $M_i(g)$, $M_i^{\beta_0}(g) = M_i(g)$ and therefore $\beta_0 = 0$. As a consequence, as both $\mathbb{E}[\epsilon_i(g)] = 0$ and $\mathbb{E}[\delta] = 0$:

$$\forall \delta \in F_0, \forall i \in \llbracket 1, p \rrbracket, \forall g \in \mathcal{X}, \quad \mathbb{E}[\epsilon_i(g) \delta] = \text{Cov}[\epsilon_i(g), \delta] = 0. \quad (14)$$

From a geometrical point of view it is equivalent to say that the inner product of the error and any vector of F_0 , such as the difference of any linear unbiased predictors of $Y_j(g')$, is null. This approach can be found for example in Aldworth [1998], section 4.5.1. page 122, in the case of ordinary Kriging on a stationary process.

- Now, let δ and δ' be any two vectors of F_0 . As a consequence of the previous result in Equation (14), we have:

$$\text{Cov}[\epsilon_i(g) + \delta, \epsilon_j(g') + \delta'] = c_{i,j}(g, g') + 0 + 0 + \text{Cov}[\delta, \delta'] \quad (15)$$

- On the other hand, using the conditional covariance formula, we have:

$$\text{Cov}[\epsilon_i(g) + \delta, \epsilon_j(g') + \delta'] = \mathbb{E}[\text{Cov}[\epsilon_i(g) + \delta, \epsilon_j(g') + \delta' \mid \underline{\mathbf{Y}}]] + \text{Cov}[\mathbb{E}[\epsilon_i(g) + \delta \mid \underline{\mathbf{Y}}], \mathbb{E}[\epsilon_j(g') + \delta' \mid \underline{\mathbf{Y}}]]$$

Given a $\underline{\mathbf{Y}}$, the random variables δ , δ' , $M_i(g)$ and $M_j(g')$ are constant, so that the first term is

$$\mathbb{E}[\text{Cov}[\epsilon_i(g) + \delta, \epsilon_j(g') + \delta' \mid \underline{\mathbf{Y}}]] = \mathbb{E}[\text{Cov}[Y_i(g), Y_j(g') \mid \underline{\mathbf{Y}}]].$$

Furthermore, we have assumed in Assumption (A) that $M_i(g) = \mathbb{E}[Y_i(g) \mid \underline{\mathbf{Y}}]$ and $M_j(g') = \mathbb{E}[Y_j(g') \mid \underline{\mathbf{Y}}]$, therefore $\mathbb{E}[\epsilon_i(g) \mid \underline{\mathbf{Y}}] = \mathbb{E}[\epsilon_j(g') \mid \underline{\mathbf{Y}}] = 0$ and:

$$\text{Cov}[\epsilon_i(g) + \delta, \epsilon_j(g') + \delta'] = \mathbb{E}[\text{Cov}[Y_i(g), Y_j(g') \mid \underline{\mathbf{Y}}]] + \text{Cov}[\delta, \delta'] \quad (16)$$

Identifying the equations (15) and (16), we get the expected result.

□

Note that the uncertainty relying on each X_g can be propagated to $\underline{\mathbf{Y}}$ through Proposition 1, and then to the prediction of \mathbf{Y}_g over each grain $g \in \mathcal{G}$ through Propositions 2, 3 or 4, using $v_i(g) = c_{i,i}(g, g)$. This answers the problem(P2) presented in the introduction.

4 Numerical illustrations

The numerical use cases provided here show that mixture Kriging can be applied to solve some **uncertainty propagation problems**. First, we use test functions. Then we use real data to show how mixture Kriging can improve existing state of the art Kriging models. In this last example, we also present an approach to optimize parameters. We have developed a mixture kriging package in R language. It also allows fitting a classical Kriging model (case where grains are restricted to points).

4.1 With simulated data: rounded inputs

A common issue for those who feed statistical models with real data is the question of data accuracy. In this example we assume that input data is rounded to the nearest unit. We observe the consequences on a classical Kriging model versus a mixture Kriging model. In the following example, visualized in Figure 3, we draw 6 observations at 0.55, 0.85, 7.20, 3.00, 3.45, 9.40 from a simulated uniform centered gaussian random field $Y(x)$ defined for $x \in [1, 10]$ (200 points). Observed x values are rounded to the nearest unit 1, 1, 7, 3, 3, 9 before feeding the models. The 3 first observations are picked to train a model which is used to make predictions. In other words, we forget the underlying simulation and now assume that $Y(x)$ is conditioned by the observations number 1, 2, 5. A gaussian kernel is used. Nugget effect and lengthscales are optimized with a particle swarm optimization algorithm (R package `hydroPSO`) predicting the output on the 3 other observed points/grains (3, 4, 6) and mean squared error is minimized.

Classical Kriging Point observations are possible, not grains. Therefore $x = 1$ and $x = 3$ are observed twice each, while $x = 7$ and $x = 9$ are observed once each. Classical Kriging (Figure 3 right) can not handle repeated observations if not adding a nugget effect for covariance matrix inversibility. Meaning that uncertainty on the input value is seen by design as an uncertainty on the output value. As a result, nugget effect is over-estimated and prediction (green line) is flattened along the random field mean value and prediction uncertainty (confidence band) is very large.

Mixture Kriging Grains $[0.5, 1.5[$ and $[2.5, 3.5[$ are observed twice each while grains $[6.5, 7.5[$ and $[8.5, 9.5[$ are observed once each. Mixture Kriging model (Figure 3 left) can handle repeated observations by design. Uncertainty on the input is a result of observing a grain on a random position. See the way grain covariances are computed from point covariance in Proposition 1. In that case, nugget effect is not overestimated, mean prediction variability is greater than in classical Kriging and prediction uncertainty is smaller.

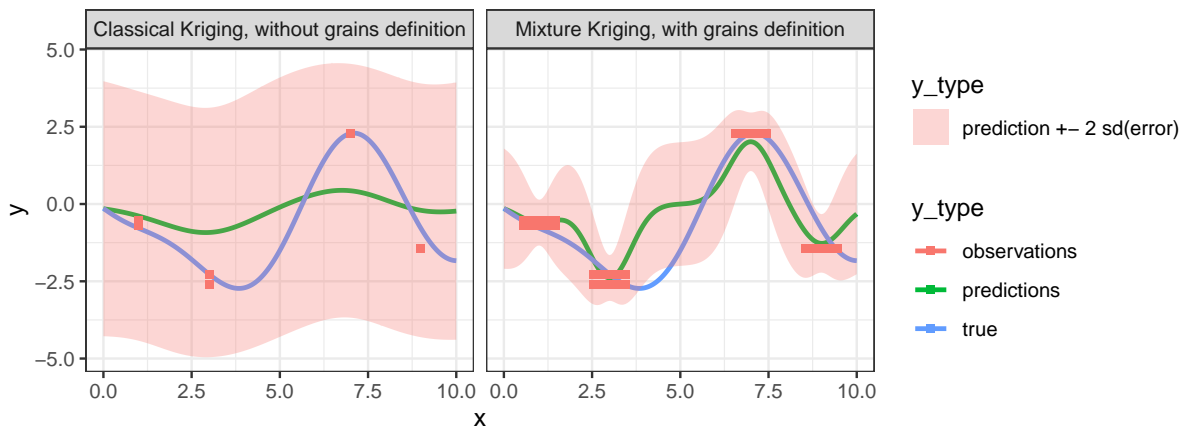


Figure 3: *Effect of rounded inputs on Kriging.* **Left and right** Blue line labelled “true” shows a simulated uniform random field of variance 5 and lengthscales 3. **Left:** Red line segments show 6 grain observations. Grains are intervals of length 1. Green line shows mixture Kriging model after optimizing mean squared error. **Right** Red dots show 6 point observations of this field for 6 values of x that are rounded to the nearest unit. Green line labelled “predictions” shows classical Kriging model after optimizing mean squared error. **Left and right** Pink ribbon shows an interval of radius twice the estimated error standard deviation.

4.2 With simulated data: Uncertainty on input data

Imagine an industry, providing some manufactured objects. The company aims at measuring some performance indicators $\mathbf{Y}(x)$ of a manufactured object, depending on a design $x \in \chi$: e.g. it can measure the lift of an aircraft wing, depending on a specific design (described by some shape parameters). But some uncertainties may rely on the design, as a constructed object has slightly different characteristics from the theoretical design. Thus, the constructed design can be considered as a random vector X_{g_x} , taking values in some tolerance set $g_x \subset \chi$ around the design $x \in \chi$ (as such it is a generalization of the previous example). As a consequence, when testing different possible designs x_1, \dots, x_k , the industry

observes $\mathbf{Y}(g_1), \dots, \mathbf{Y}(g_k)$. Testing a new possible design can be costly, so that finding the best design requires to get a prediction of both the expectation and the variance of $\mathbf{Y}(x)$, for each considered new possible design $x \in \chi$. In this setting, for the sake of simplicity, we assume that $\mathbf{Y}(x)$ is conditioned by $\mathbf{Y}(x) = \mathbf{y}_{\text{reference}}(x)$ at observation points x_1, \dots, x_k . Starting from given covariances between points, we have built the grain covariances of Proposition 1. In a second step, we have built the predictor presented in Proposition 2. The results appear in Figures 4, 5 and 6. The setting details are given in figures captions.

In Figure 4 (left), grains are restricted to singletons therefore classical Kriging results are retrieved. When grains are not restricted to singletons in Figure 4 (right), the uncertainty on the input is propagated to the output, so that both Kriging mean and confidence intervals are affected. In particular, the Kriging mean is no more interpolating.

In Figure 5, we illustrate the situation of varying grain sizes and resulting effect on uncertainty. It appears, as one would expect, that the wider the grains, the higher the uncertainty. One can also check that the centroid of the grain is not sufficient to model the behavior of the response: the latter depends on the size of the grain. Reducing a grain to its centroid would end up in a very different model. The reader may refer to Goovaerts [2006] to find a fully developed model where grains are restricted to their centroid (area-to-area Kriging).

In Figure 6, we illustrate the possibility of integrating aggregated data with pointwise observations. We suppose that we know the expectation and variance of y on some given grains. We feed the model not only with point observations but also with grains observations which values are the means of $y_{\text{reference}}$ on the grains. The points covariance matrix \mathbf{K} is completed with a diagonal block containing grains variances. We observe a great reduction of prediction uncertainty in the second case: the confidence band on the right plot is thinner than the left one. The Kriging mean is also modified near $x = 0$.

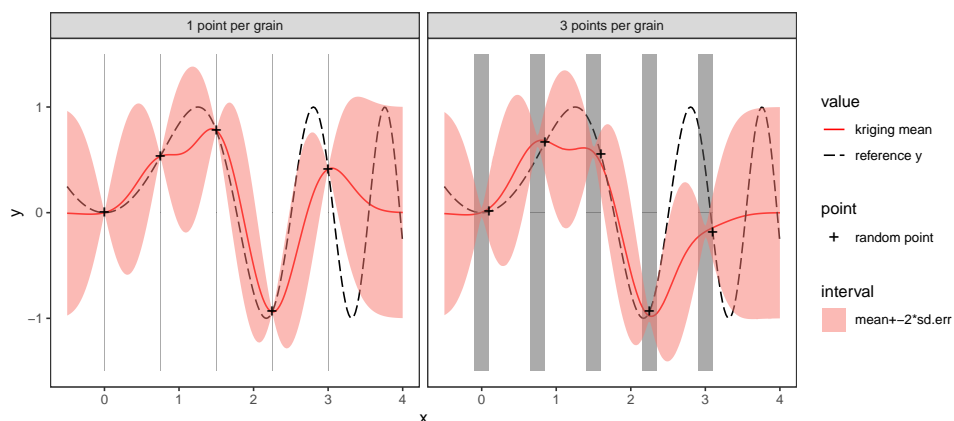


Figure 4: *Comparison of classical simple Kriging and mixture simple Kriging.* **Left:** Reference function $y_{\text{reference}}$ (dashed line) is observed at 5 points without uncertainty on the value of x . We find the usual results on simple Kriging, in particular predictions (solid line) are interpolating. **Right:** For each observation, x is an unknown random value among 3 possible values $x_i, x_i + r, x_i - r$. Predictions are not interpolating any more. We also observe that the estimated standard deviation of the error on y is slightly increased between the grains. **Details:** In this example, $y_{\text{reference}}(x) = \sin(x^2)$, covariance is Gaussian with $\sigma = 0.5$ and $\theta = 0.3$, X_g is discrete uniform on each grain, $r = 0.1$.

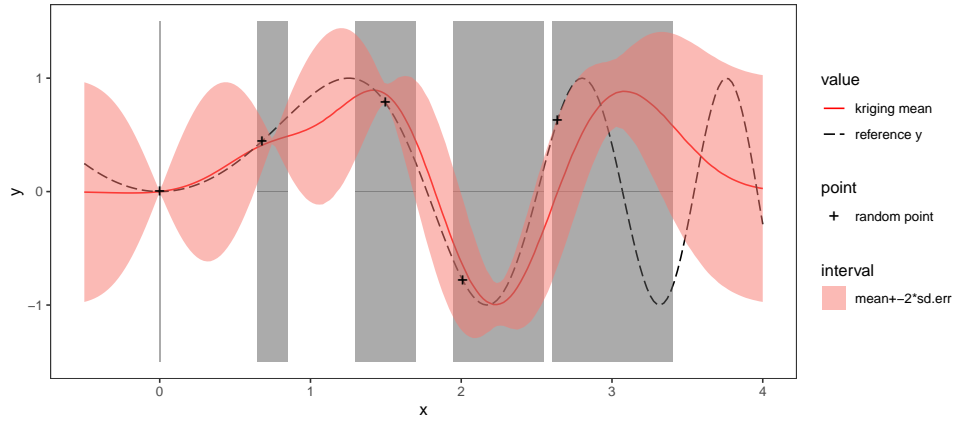


Figure 5: *Varying grain sizes and effect on uncertainty.* Reference function y (solid line) is observed at 5 grains of different sizes. We observe that predictions uncertainty is growing when the grain size is growing. **Details:** In this example, $y_{\text{reference}}(x) = \sin(x^2)$, covariance is Gaussian with $\sigma = 0.5$ and $\theta = 0.3$, X_g is continuous uniform on each grain.

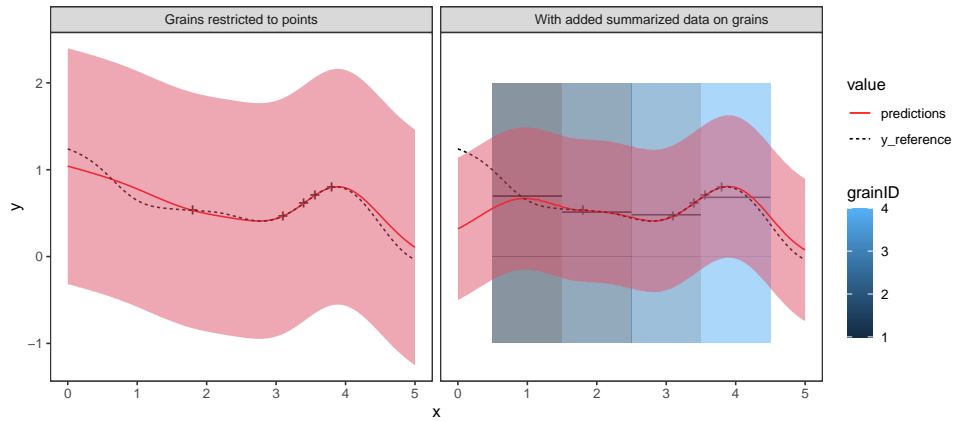


Figure 6: *Effect of knowing aggregated information on a grain.* **Left:** Ordinary Kriging in its classical form, with grains restricted to points. **Right:** In addition to the observations on points, we assume that we know expectation and variance of y over 4 grains. We integrate this data both as new observations evaluated to expectation value on the grains but also with added variance on the diagonal of the covariance matrix \mathbf{K} . **Details:** Reference function is a simulation of a Gaussian process with Gaussian covariance ($\sigma = 0.4$, $\theta = 0.6$), expectation of y being $1 - 0.2x$, X_g is discrete uniform on the grains.

4.3 With real data: building energy efficiency

Due to climate changes and energy crisis, it is a major concern for governments be it national or regional, to assess the buildings energy efficiency. In France, URBS enterprise has to address this issue for its customers. Given the fact that only 15% to 20% of housing buildings have been diagnosed for energy efficiency in the last decade, it has to find new approaches to make large scale predictions of buildings energy efficiency. It is well known in literature and by experts that buildings age is a major source of information to assess energy efficiency (see for instance Ballarini et al. [2017]). Among several important variables identified in the literature, we focus on the age for the sake of illustration and readability. However we must keep in mind first that this variable “year of construction” defines an age rounded to the nearest unit and second that a building may consist of dwellings that have been built at different times. The rounding effect is particularly important for very recent buildings, built in a period of constant and fast legal changes. The second phenomenon is more important for old buildings which may have been extended at some point in their life. Even recent works (see for instance Schetelat et al. [2020]) ignore both facts resulting in information loss.

We try to improve the use of construction year to predict buildings energy efficiency which is measured as an energy consumption given in $kWh/m^2/year$ (kilowatt-hour per square meter and per year). In a first model, say M_1 , we associate each building in Auvergne-Rhône-Alpes (AURA – one of France’s regions) with a single construction year. Common use is to pick the most recent one but it could be any statistic. A construction year difference of 10 years in the 21st century may have a high impact on the energy efficiency while it has almost no impact in the 19th century. For this reason “year of construction” values have been shrunked to $[0, 1]$ using ranks, then normalized on \mathbb{R} using standard normal quantiles. Same transformation is made on energy efficiency values for similar reasons. We model energy efficiency as a random field on \mathbb{R} (normalized year of construction). This field is observed on a discrete set of values (year of construction rounded to the nearest unit and normalized). We fit a classical simple Kriging model that is optimized with R package `DiceKriging` which is a commonly used package for standard Kriging. We assume that covariance kernel is Matérn 3/2 and estimate associated variance, lengthscale and nugget effect.

In a second model, say M_2 , we associate each building with a set of intervals of length 1 hereby defining a granularity. Each grain of this granularity consists of one or more intervals of length 1. If a grain has 2 or more intervals, it is called a compound grain. A random position is defined on those grains so that it is uniform on each interval and when a grain contains multiple intervals, the probability to be on one interval or the other is proportional to the associated surface area of dwellings in the building. Energy efficiencies that have been measured by technicians are seen as observations of these grains. Construction year and energy efficiency are normalized as described above. We fit a mixture Kriging model that is optimized on a grid of parameters values for variance, lengthscale and nugget effect. We assume that covariance kernel is Matérn 3/2.

From the energy efficiency, one can derive a label ranging from A to G. Our end goal is to improve the balanced accuracy of those labels prediction. The balanced accuracy is the mean value of the percentage of good predictions for each label.

Validation process We pick 500 grains randomly among the observed grains. We collect all 1570 observations on those grains. This is the training set for M_1 . The same observations are used for training M_0 . Optimization is conducted with cross validation where split is defined grouping all observations of a grain in a same subgroup. Once optimized, the resulting model is used to predict the output on a validation set built similarly with 250 new grains (690 observations).

Characteristics of M_0 and M_1 are detailed in Table 1. Both models have a similar estimated value for variance. As expected from simulations, the classical model M_0 overestimates the nugget effect, resulting in a large rmse as compared to M_1 . Validation balanced accuracy is slightly improve when using mixtures. And most of the improvement is concentrated on compound grains. It appears that mixture Kriging is able to value information given by compound grains when this same information is partially lost in M_0 . Note that since we use only one input variable in this illustration, we can not expect a much greater balanced accuracy. We are currently working on a detailed model with more variables and different kind of data such as anonymized census data.

		M_1	M_2
Optimized parameters	variance	0.258	0.260
	lengthscale	0.031	0.262
	nugget effect	0.717	0.010
Validation performances	energy efficiency rmse ↓	164	89
	balanced accuracy ↑	17.9%	19.3%
	balanced accuracy restricted to compound grains ↑	17.5%	23.4%

Table 1: Classical simple Kriging versus mixture simple Kriging performances on real data. ↓: the lower the better, ↑: the greater the better.

5 Conclusion

In this work we have introduced a new setup based on granularities, able to model input data uncertainty with mixture distributions. We have shown that it is possible to derive from this data model a best linear unbiased predictor which is also able to assess the prediction error. The conditions to apply this mixture Kriging model is to be able to compute the covariance between output variables at 2 different points of the territory. We have treated here the case where the output variable has constant expectation over the territory. We have presented some potential use cases of this model. We have applied this model to real data with success to catch extra information that is available in raw data but partially lost in state of the art current models. We plan to apply it on larger datasets with multiple variables. We also study the possibility to develop a similar model with Universal Kriging.

Acknowledgements

The authors acknowledge support from the URBS entreprise, www.urbs.fr. They thank in particular Maximilien Brossard and Jonathan Villot and for careful reading and constructive comments.

This research is jointly funded by Mines Saint-Etienne graduate engineering school and research institute (<https://www.mines-stetienne.fr/en/>), URBS entreprise and French National Agency for Research and Technology (<https://www.anrt.asso.fr/fr>).

References

- Walter Jeremy Koch Aldworth. *Spatial prediction, spatial sampling, and measurement error*. Doctor of Philosophy, Iowa State University, Digital Repository, Ames, 1998. URL <https://lib.dr.iastate.edu/rtd/11842/>. Pages: 6510332.
- Ilaria Ballarini, Vincenzo Corrado, Francesco Madonna, Simona Paduos, and Franco Ravasio. Energy refurbishment of the Italian residential building stock: energy and cost analysis through the application of the building typology. *Energy Policy*, 105:148–160, June 2017. ISSN 0301-4215. doi: 10.1016/j.enpol.2017.02.026. URL <http://www.sciencedirect.com/science/article/pii/S0301421517301015>.
- Álvaro Briz-Redón. A Bayesian shared-effects modeling framework to quantify the modifiable areal unit problem. *Spatial Statistics*, 51:100689, October 2022. ISSN 2211-6753. doi: 10.1016/j.spasta.2022.100689. URL <https://www.sciencedirect.com/science/article/pii/S2211675322000537>.
- Alan E. Gelfand, Li Zhu, and Bradley P. Carlin. On the change of support problem for spatio-temporal data. *Biostatistics*, 2(1):31–45, March 2001. ISSN 1465-4644. doi: 10.1093/biostatistics/2.1.31. URL <https://academic.oup.com/biostatistics/article/2/1/31/238352>. Publisher: Oxford Academic.
- Pierre Goovaerts. Geostatistical analysis of disease data: accounting for spatial support and population density in the isopleth mapping of cancer mortality risk using area-to-point Poisson kriging. *International Journal of Health Geographics*, 5:52, November 2006. ISSN 1476-072X. doi: 10.1186/1476-072X-5-52. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1697809/>.

- Pierre Goovaerts. Kriging and Semivariogram Deconvolution in the Presence of Irregular Geographical Units. *Mathematical Geology*, 40(1):101–128, 2008. ISSN 0882-8121.
- Carol Gotway and Linda Young. Combining Incompatible Spatial Data. *Journal of the American Statistical Association*, 97:632–648, February 2002. doi: 10.1198/016214502760047140.
- Yan Jin, Yong Ge, Jianghao Wang, Gerard Heuvelink, and Le Wang. Geographically Weighted Area-to-Point Regression Kriging for Spatial Downscaling in Remote Sensing. *Remote Sensing*, 10:579, April 2018. doi: 10.3390/rs10040579.
- Ruth Kerry, Pierre Goovaerts, Izak P.J. Smit, and Ben R. Ingram. A comparison of multiple indicator kriging and area-to-point Poisson kriging for mapping patterns of herbivore species abundance in Kruger National Park, South Africa. *International journal of geographical information science : IJGIS*, 27(1):47–67, 2013. ISSN 1365-8816. doi: 10.1080/13658816.2012.663917. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4341904/>.
- Phaedon Kyriakidis. A Geostatistical Framework For Area-To-Point Spatial Interpolation. *Geographical Analysis*, 36, August 2004. doi: 10.1353/geo.2004.0009.
- Changjiang Li, Zhiming Lu, Tuhua Ma, and Xingsheng Zhu. A simple kriging method incorporating multiscale measurements in geochemical survey. *Journal of Geochemical Exploration*, 101(2):147–154, May 2009. ISSN 0375-6742. doi: 10.1016/j.gexplo.2008.06.003. URL <http://www.sciencedirect.com/science/article/pii/S0375674208000666>.
- Carlos Lima, Hyun-Han Kwon, and Yong-Tak Kim. A Bayesian Kriging Model Applied for Spatial Downscaling of Daily Rainfall from GCMs. *Journal of Hydrology*, 597, February 2021. doi: 10.1016/j.jhydrol.2021.126095.
- Yu-Pin Lin, Bai-You Cheng, Guey-Shin Shyu, and Tsun-Kuo Chang. Combining a finite mixture distribution model with indicator kriging to delineate and map the spatial patterns of soil heavy metal pollution in Chunghua County, central Taiwan. *Environmental Pollution*, 158(1):235–244, January 2010. ISSN 0269-7491. doi: 10.1016/j.envpol.2009.07.015. URL <https://www.sciencedirect.com/science/article/pii/S0269749109003534>.
- Oswaldo José Ribeiro Pereira, Adolpho José Melfi, Célia Regina Montes, and Yves Lucas. Downscaling of ASTER Thermal Images Based on Geographically Weighted Regression Kriging. *Remote Sensing*, 10(4):633, April 2018. doi: 10.3390/rs10040633. URL <https://www.mdpi.com/2072-4292/10/4/633>. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.
- Laura Poggio and Alessandro Gimona. Downscaling and correction of regional climate models outputs with a hybrid geostatistical approach. *Spatial Statistics*, 14:4–21, November 2015. ISSN 2211-6753. doi: 10.1016/j.spasta.2015.04.006. URL <https://www.sciencedirect.com/science/article/pii/S2211675315000305>.
- Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass, 2006. ISBN 978-0-262-18253-9. OCLC: ocm61285753.
- Pascal Schetelat, Lucie Lefort, and Nicolas Delgado. Urban data imputation using multi-output multi-class classification. *Building to Buildings: Urban and Community Energy Modelling*, November 2020.
- Phuong Truong and Gerard Heuvelink. Bayesian Area-to-Point Kriging using Expert Knowledge as Informative Priors. *International Journal of Applied Earth Observation and Geoinformation*, 30:2291, April 2013. doi: 10.1016/j.jag.2014.01.019.
- Qunming Wang, Wenzhong Shi, and Peter M. Atkinson. Area-to-point regression kriging for pan-sharpening. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114:151–165, April 2016. ISSN 0924-2716. doi: 10.1016/j.isprsjprs.2016.02.006. URL <http://adsabs.harvard.edu/abs/2016JPRS..114..151W>.
- Christopher Williams and Carl Rasmussen. Gaussian Processes for Regression. *Advances in Neural Information Processing Systems 8*, 8, March 1996.

Simon N. Wood. *Generalized Additive Models: An Introduction with R, Second Edition*. CRC Press, May 2017. ISBN 978-1-4987-2834-8.

E.-H. Yoo and P. C. Kyriakidis. Area-to-point Kriging with inequality-type data. *Journal of Geographical Systems*, 8(4):357–390, November 2006. ISSN 1435-5930, 1435-5949. doi: 10.1007/s10109-006-0036-7. URL <http://link.springer.com/10.1007/s10109-006-0036-7>.

Xiaohu Zhang, Wenjun Zuo, Shengli Zhao, Li Jiang, Linhai Chen, and Yan Zhu. Uncertainty in Upscaling In Situ Soil Moisture Observations to Multiscale Pixel Estimations with Kriging at the Field Level. *ISPRS International Journal of Geo-Information*, 7(1):33, January 2018. doi: 10.3390/ijgi7010033. URL <https://www.mdpi.com/2220-9964/7/1/33>. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

A Appendix: Operations on granularities

In the course of our research, we started studying some granularities available in our databases and their relations/classifications: e.g. what is the relation between the set of land plots and the set of census tracts? We also thought about ways to build non-overlapping granularities from existing granularities. This led us to the definitions of the following concepts.

Definition 4 (Non-overlapping granularity). *A granularity \mathcal{G} is said to be **non-overlapping** when all intersections of grains are empty: $\forall g, g' \in \mathcal{G}, g \cap g' = \emptyset$.*

Definition 5 (Granularity order). *The **granularity order** $\mathcal{G} \leq \mathcal{H}$, or equivalently $\mathcal{H} \geq \mathcal{G}$, holds for two granularities \mathcal{G} and \mathcal{H} under the following condition:*

$$\mathcal{G} \leq \mathcal{H} \Leftrightarrow \forall g \in \mathcal{G}, \left\{ \begin{array}{l} g \in \bigcup_{h \in \mathcal{H}} h \\ \text{and } \forall h \in \mathcal{H}, g \cap h \in \{\emptyset, g\} \end{array} \right.$$

\mathcal{G} is said to be *thinner* than \mathcal{H} , or equivalently \mathcal{H} *coarser* than \mathcal{G} . In particular, $\mathcal{G} \leq \mathcal{H}$ implies that any grain of \mathcal{G} is a subset of at least one grain in \mathcal{H} , but it also implies that a grain of \mathcal{G} does not partly overlap a grain of \mathcal{H} .

Relation \leq is transitive on the set of granularities defined on χ . It defines a partial order on this set.

Proposition 5 (Non-overlapping granularities). *Define an **insertion operator** \oplus , for any non-overlapping granularity \mathcal{G} and any grain h by:*

$$\mathcal{G} \oplus \{h\} := \left\{ g_0 : g_0 \neq \emptyset \text{ and } g_0 \in \{g \cap h : g \in \mathcal{G}\} \cup \{g \setminus h : g \in \mathcal{G}\} \cup \left\{ h \setminus \bigcup_{g \in \mathcal{G}} g \right\} \right\}.$$

This operator \oplus adds a partition of the grain h to the non-overlapping granularity \mathcal{G} , while ensuring that $\mathcal{G} \oplus \{h\}$ is non-overlapping and has the same union of grains as $h \cup \bigcup_{g \in \mathcal{G}} g$.

Then we have:

(i) For any non-overlapping granularity \mathcal{G} and grain h , the resulting granularity is thinner than $\mathcal{G} \cup \{h\}$:

$$\mathcal{G} \oplus \{h\} \leq \mathcal{G} \cup \{h\}.$$

(ii) For any non-overlapping granularity \mathcal{G} and grains h, h' , the insertion order does not matter:

$$(\mathcal{G} \oplus \{h\}) \oplus \{h'\} = (\mathcal{G} \oplus \{h'\}) \oplus \{h\}.$$

(iii) Among the granularities that are thinner than a finite granularity $\mathcal{G} = \{g_1, \dots, g_n\}$, there is a unique **maximal non-overlapping granularity** \mathcal{G}^\oplus and we can construct it iteratively with the insertion operator.

$$\mathcal{G}^\oplus := \{g_1\} \oplus \dots \oplus \{g_n\}. \quad (17)$$

This granularity is a non-overlapping granularity such that $\mathcal{G}^\oplus \leq \mathcal{G}$, and it is maximal, in the sense that any other non-overlapping \mathcal{G}' that is thinner than \mathcal{G} is also thinner than \mathcal{G}^\oplus : $\mathcal{G}' \leq \mathcal{G} \Rightarrow \mathcal{G}' \leq \mathcal{G}^\oplus$.

Proof. • Let us prove the item (i)

Let us prove that $\mathcal{G} \oplus \{h\} \leq \mathcal{G} \cup \{h\}$. Let $g_+ \in \mathcal{G} \oplus \{h\}$ and $g' \in \mathcal{G} \cup \{h\}$. It is clear by construction that $g_+ \in \bigcup_{g \in \mathcal{G} \cup \{h\}} g$. Moreover:

$$g_+ = g \cap h \text{ or } g_+ = g \setminus h \text{ or } g_+ = h \setminus \bigcup_{g \in \mathcal{G}} g \quad \text{AND} \quad g' \in \mathcal{G} \text{ or } g' = h$$

One can prove that in all 6 different combined cases, either $g_+ \cap g' = g_+$ or $g_+ \cap g' = \emptyset$.

As a consequence, $\mathcal{G} \oplus \{h\} \leq \mathcal{G} \cup \{h\}$.

- Let us prove the item (ii).
Let $g_2 \in (\mathcal{G} \oplus \{h\}) \oplus \{h'\}$ then:

$$(A) \exists g_1 \in \mathcal{G} \oplus \{h\}, g_2 = g_1 \cap h' \quad \text{or} \quad (B) \exists g_1 \in \mathcal{G} \oplus \{h\}, g_2 = g_1 \setminus h' \quad \text{or} \quad (C) g_2 = h' \setminus \bigcap_{g \in \mathcal{G} \oplus \{h\}} g$$

Let $g_1 \in \mathcal{G} \oplus \{h\}$ then:

$$(a) \exists g_0 \in \mathcal{G}, g_1 = g_0 \cap h \quad \text{or} \quad (b) \exists g_0 \in \mathcal{G}, g_1 = g_0 \setminus h \quad \text{or} \quad (c) g_1 = h \setminus \bigcup_{g \in \mathcal{G}} g$$

$$\begin{array}{lll} (Aa) & g_2 = g_0 \cap h \cap h' & = g_0 \cap h' \cap h & \in (\mathcal{G} \oplus \{h'\}) \oplus \{h\}, \text{ see case (Aa)} \\ (Ab) & g_2 = (g_0 \setminus h) \cap h' & = (g_0 \cap h') \setminus h & \in (\mathcal{G} \oplus \{h'\}) \oplus \{h\}, \text{ see case (Ba)} \\ (Ac) & g_2 = (h \setminus \bigcup_{g \in \mathcal{G}} g) \cap h' & = (h' \setminus \bigcup_{g \in \mathcal{G}} g) \cap h & \in (\mathcal{G} \oplus \{h'\}) \oplus \{h\}, \text{ see case (Ac)} \\ (Ba) & g_2 = (g_0 \cap h) \setminus h' & = (g_0 \setminus h') \cap h & \in (\mathcal{G} \oplus \{h'\}) \oplus \{h\}, \text{ see case (Ab)} \\ (Bb) & g_2 = (g_0 \setminus h) \setminus h' & = (g_0 \setminus h') \setminus h & \in (\mathcal{G} \oplus \{h'\}) \oplus \{h\}, \text{ see case (Bb)} \\ (Bc) & g_2 = (h \setminus \bigcup_{g \in \mathcal{G}} g) \setminus h' & = h \setminus \bigcup_{g \in \mathcal{G} \oplus \{h'\}} g & \in (\mathcal{G} \oplus \{h'\}) \oplus \{h\}, \text{ see case (C)} \\ (C) & g_2 = h' \setminus \bigcup_{g \in \mathcal{G} \oplus \{h\}} g & = (h' \setminus \bigcup_{g \in \mathcal{G}} g) \setminus h & \in (\mathcal{G} \oplus \{h'\}) \oplus \{h\}, \text{ see case (Bc)} \end{array}$$

For cases (Bc) and (C), we used the fact that $\bigcup_{g \in \mathcal{G} \oplus \{h\}} g = h \cup \bigcup_{g \in \mathcal{G}} g$.

- Let us prove the item (iii)
Note that due to item (ii), \mathcal{G}^\oplus does not depend on the indexing order of the grains composing \mathcal{G} .
Moreover, due to item (i), $\{g_1\} \oplus \{g_2\} \leq \{g_1, g_2\}$ and by recurrence, $\mathcal{G}^\oplus \leq \mathcal{G}$.
Now let us prove that for any non-overlapping granularity \mathcal{H} , any granularity \mathcal{G} , any grain g_0 :

$$\mathcal{G} \leq \mathcal{H} \cup \{g_0\} \Rightarrow \mathcal{G} \leq \mathcal{H} \oplus \{g_0\}$$

Suppose $\mathcal{G} \leq \mathcal{H} \cup \{g_0\}$. Let $g \in \mathcal{G}$ and $g_+ \in \mathcal{H} \oplus \{g_0\}$, taking into account that \mathcal{H} is non-overlapping:

$$\begin{array}{lll} (A) \exists h \in \mathcal{H} : g \subset h \cap g_0 & \text{or} & (B) \exists h \in \mathcal{H} : g \subset h \setminus g_0 & \text{or} & (C) g \subset g_0 \setminus \bigcup_{h \in \mathcal{H}} h \\ \text{and (a) } \exists h' \in \mathcal{H} : g_+ = h' \cap g_0 & \text{or} & (b) \exists h' \in \mathcal{H} : g_+ = h' \setminus g_0 & \text{or} & (c) g_+ = g_0 \setminus \bigcup_{h \in \mathcal{H}} h \end{array}$$

In cases Ab, Ac, Ba, Bc, Ca, Cb, we have $g \cap g_+ = \emptyset$. In cases Aa and Bb, if $h = h'$ then $g \cap g_+ = g$, otherwise $g \cap g_+ = \emptyset$. In case Cc, $g \cap g_+ = g$. Therefore in either case, $g \cap g_+ \in \{g, \emptyset\}$ and $\mathcal{G} \leq \mathcal{H} \oplus \{g_0\}$. □

When a non-overlapping granularity is needed, one can thus use Proposition 5 and build \mathcal{G}^\oplus directly from any finite granularity \mathcal{G} , possibly overlapping. However, we will see in the rest of the paper that the proposed model is also suited for overlapping granularities.

When two data sources are available, relying on two granularities \mathcal{G} and \mathcal{H} it can also be convenient to define $\mathcal{G} \oplus \mathcal{H} := (\mathcal{G} \cup \mathcal{H})^\oplus$ to get a non-overlapping resulting granularity allowing to work with both data sources. As an example, if an information is given at the level of a grid reference system \mathcal{G} , and also at a level of urban areas \mathcal{H} , it may be convenient to build all intersection areas by this way. The Proposition 5 gives a simple way to do so, even in more complicated situations where both \mathcal{G} and \mathcal{H} are overlapping granularities.

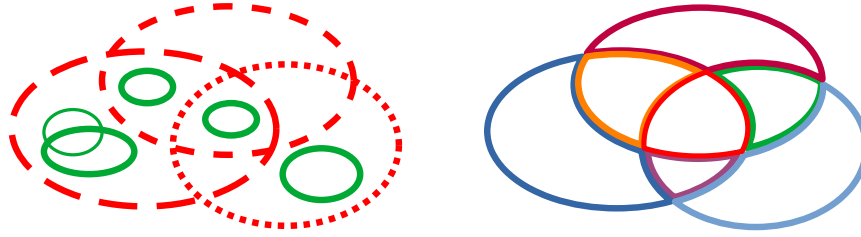


Figure 7: *Thinner granularity and maximal thinner non-overlapping granularity.* **Left:** The granularity comprising the 5 green grains (solid lines) is thinner than the granularity comprising the 3 red grains (dashed lines). **Right:** The granularity comprising 7 non overlapping green grains is the maximal non-overlapping granularity that is thinner than the red granularity on the left.

Contents

1	Introduction	1
2	Spatial model	4
2.1	Inputs	4
2.2	Outputs	5
3	Prediction	8
4	Numerical illustrations	13
4.1	With simulated data: rounded inputs	14
4.2	With simulated data: Uncertainty on input data	14
4.3	With real data: building energy efficiency	17
5	Conclusion	18
A	Appendix: Operations on granularities	21