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. 2021. hal-03276127v1

**HAL Id: hal-03276127**

**https://hal.science/hal-03276127v1**

Preprint submitted on 1 Jul 2021 (v1), 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 30th June, 2021‡

**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 specific locations, which creates specific problems of multiscale areal data. Second, some uncertainties rely both on the input locations and on measured quantities at these locations, which creates specific 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 covariances can be derived from this specific setting. A numerical illustration on simulated data is given.

***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 (with, say, $\chi = \mathbb{R}^2$), or for instance an industrial design summarized by some parameters (with $\chi = \mathbb{R}^d$, $d \in \mathbb{N}^*$). 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 $\chi$, 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 $y(.)$ 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 a possible 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 $Y(.)$ is Gaussian with mean zero, the simple Kriging predictor ends up in a (posterior) conditional Gaussian process having conditional Kriging mean and covariance, for any $x, x' \in \chi$:

$$\left\{ \begin{array}{lcl} \mathbb{E}\left[Y(x)|Y(x_i) = y(x_i),\, i = 1, \dots, n\right] & = & \mathbf{h}(x)^\top \mathbf{K}^{-1}\mathbf{y} \\ \mathrm{Cov}\left[Y(x), Y(x')|Y(x_i) = y(x_i),\, i = 1, \dots, n\right] & = & k(x, x') - \mathbf{h}(x)^\top \mathbf{K}^{-1}\mathbf{h}(x') \end{array} \right. \tag{1}$$

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:

$$\left\{ \begin{array}{lcl} k(x, x') & = & \mathrm{Cov}\left[Y(x), Y(x')\right] \in \mathbb{R} \\ \mathbf{h}(x) & = & (\mathrm{Cov}\left[Y(x_i), Y(x)\right])_{i=1,\dots,n} \in \mathbb{R}^{n\times 1} \\ \mathbf{K} & = & (\mathrm{Cov}\left[Y(x_i), Y(x_j)\right])_{i,j\in\{1,\dots,n\}} \in \mathbb{R}^{n\times n} \end{array} \right.$$

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 outputs can be missing. To achieve this goal, we extend classical Kriging techniques to our new settings.

*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, Departement GMI, Espace Fauriel, 29 rue Ponchardier, F - 42023 Saint-Etienne, France. didier.rulliere@emse.fr

‡version of this document: Wednesday 30th June, 2021, 16:42.

**Granular data** In this paper, we 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 not only for points of the input space but also 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 form of dependence between outputs based on the relative positions of the associated inputs.

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 data is usual for satellite images for instance, in which case data is 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 consists in 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 creates uncertainty propagation problems. 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 highly convex function on an average salary will not only result in a loss of randomness, but also in a convexity bias as discussed and illustrated below in Figure 2.

(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).

**Literature** There exists a vast literature dealing with the prediction 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 (such as 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. On the one hand, average is seen as positive as its dispersion is reduced as compared to point values; but on the other hand, it is clearly noted that averaging induces some problems, as we detail later on.

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 usage of 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). 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.

But it is also 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 complete 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...). In this paper, we make no difference in nature between a grain containing a continuous set of points or a discrete one.

**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 asumption 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, as we detail later on, 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 area. 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:*

- *The initial* ***territory*** *is a subset $\chi$ 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 $\chi$ can be the set of all $(latitude, longitude)$ corresponding to a country, in that case $d = 2$ and $\chi \subset \mathbb{R}^2$. An grain may correspond, for example, to a specific city, to a specific land register portion, or to a specific building. Previous Kriging models refer to blocks or areas for sets of points that are disjoints and those authors are note interested in the family itself (see for instance Kyriakidis [2004]). We develop hereafter 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 registered lands, 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 and their ordering). *In a given territory $\chi$, granularities are defined and ordered as follows:*

- *A* ***granularity*** *$\mathcal{G} = \{g_1, g_2, ...\}$ is a finite set of grains, not necessarily pairwise distinct. The granularity is said to be* ***non-overlapping*** *when all intersections of grains are empty: $\forall g, g' \in \mathcal{G},\ g \cap g' = \emptyset$.*
- *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:*

$$\mathcal{G} \leq \mathcal{H} \quad \Leftrightarrow \quad \forall g \in \mathcal{G}, \begin{cases} g \in \bigcup_{h \in \mathcal{H}} h \\ \text{and} \quad \forall h \in \mathcal{H},\ g \cap h \in \{\emptyset, g\} \end{cases}$$

*$\mathcal{G}$ 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 $\mathcal{H}$. One can check that $\mathcal{G} \leq \mathcal{H}$ implies furthermore $\mathcal{G} \leq \mathcal{H} \cup \mathcal{H}'$ for any granularity $\mathcal{H}'$.*

*Relation $\leq$ is transitive on the set of all granularities defined on $\chi$.*

**Proposition 1** (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 $\mathcal{G} \cup \{h\}$:*

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

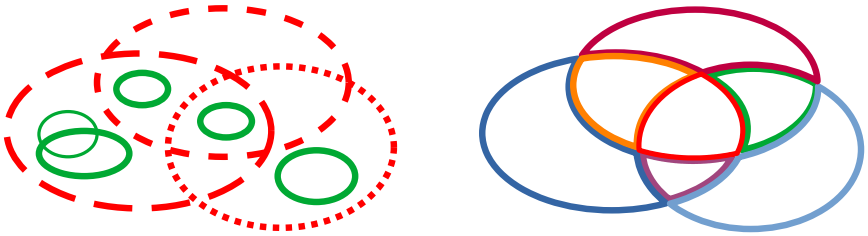

Figure 1: *Thinner granularity and maximal thinner non-overlapping granularity.* **Left:** The granularity comprising the 5 green grains (solid lines) is thinner than the granularity comprising the 3 red grains (dashed lines). **Right:** The granularity comprising 7 non overlapping green grains is the maximal non-overlapping granularity that is thinner than the red granularity on the left.

*(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 $\mathcal{G} = \{g_1, \ldots, g_n\}$, there is a unique* ***maximal non-overlapping granularity*** *$\mathcal{G}^{\oplus}$ and we can construct it iteratively with the insertion operator.*

$$\mathcal{G}^{\oplus} := \{g_1\} \oplus \ldots \oplus \{g_n\}\,. \tag{2}$$

*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.* The proof is given in Appendix □

When a non-overlapping granularity is needed, one can thus use Proposition 1 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 1 gives a simple way to do so, even in more complicated situations where both $\mathcal{G}$ and $\mathcal{H}$ are overlapping granularities.

## 2.2 Outputs

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

At each input location $x \in \chi$, the outputs may represent, say, the insulation level of a building, the households income, the square meter price, the building construction date... 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 $\chi$ denoted:*

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

- *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

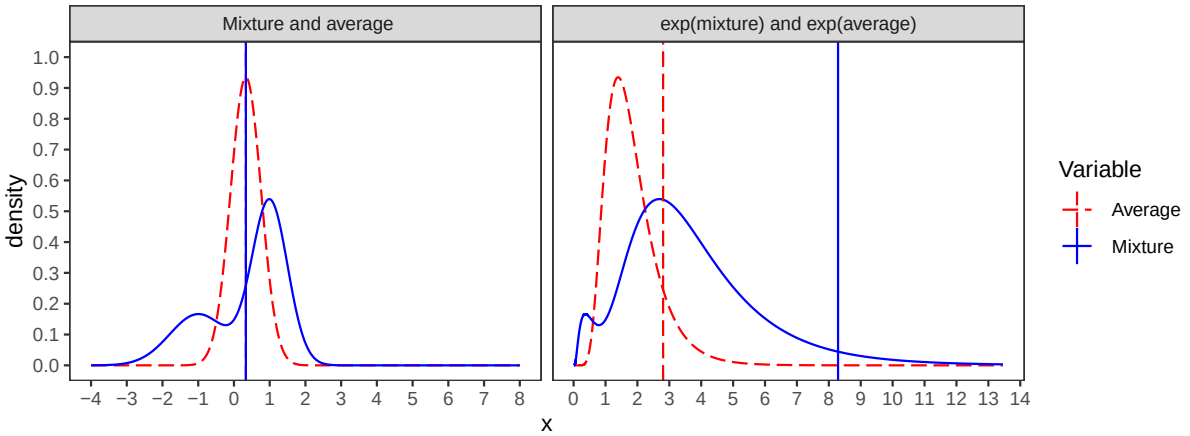


Figure 2: *Comparison of mixture and average.* **Left:** Density of a mixture $Y_{\text{mixture}}$ picking $Y_a$ with probability $\frac{1}{3}$ and picking $Y_b$ with probability $\frac{2}{3}$ (blue plain line), and density of the average $Y_{\text{average}} = \frac{1}{3}Y_a + \frac{2}{3}Y_b$ (red dashed 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

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 $\mathbb{E}\left[h(\mathbf{Y}(g))\right] \neq h\left(\mathbb{E}\left[\mathbf{Y}(g)\right]\right)$. 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. Think about life-expectancy, 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 would expect a certain stability of our process even if we apply some functions on the indicators.

- **Effects on dispersion** Let us assume that we have a grain $g = \{x_i : i \in [\![1, n]\!]\}$, $p = 1$ and consider mutually independent r.v. $\forall i$, $\mathbf{Y}(x_i) \sim \mathcal{N}(\mu, \sigma^2)$ for given $\mu$ and $\sigma$. 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 $\forall i \in [\![1, n]\!]$, $\mathbf{Y}(x_i) \sim \mathcal{U}(a, b)$ (uniform distribution on $[a, b]$) for given $a$, $b \geqslant 0$. 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, $h\left(\bar{\mathbf{Y}}(g)\right)$ depends upon $n$ and quickly converges towards $h\left(\frac{a+b}{2}\right) = \frac{a^2+2ab+b^2}{4}$. However, $\mathbf{Y}(X_g) \sim \mathcal{U}(a, b)$ does not depend on $n$ and $\mathbb{E}\left[h(\mathbf{Y}(X_g))\right] = \frac{a^2+ab+b^2}{3}$, resulting in a bias:

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

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

The Figure 2 illustrates the difference between a mixture and an average of Gaussian random variables: even when both have identical means (Figure 2, left), the mixture is no longer Gaussian and exhibits a higher variance. 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 2, 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 assumed 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 2** (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 $i \in [\![1,p]\!]$ , assuming that for all $x \in g$ we know $\mu_i(x) := \mathbb{E}[Y_i(x)]$, we have :*

$$\mu_i(g) := \mathbb{E}[Y_i(g)] = \mathbb{E}[\mu_i(X_g)] \tag{3}$$

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

$$k_{i,j}(g,g') := \mathrm{Cov}\left[Y_i(g), Y_j(g')\right] = \mathbb{E}[k_{i,j}(X_g, X_{g'})] + \mathrm{Cov}[\mu_i(X_g), \mu_j(X_{g'})] \tag{4}$$

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

*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 $X_g$). □

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 (3) and (4) can be computed easily, for all $g, g' \subset \mathcal{G}$.

Note that $\mathrm{Cov}[\mu_i(X_g), \mu_j(X_{g'})] = 0$ 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 $X_g$, $i \in [\![1,p]\!]$. 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*

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

*However*

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

*because the independence assumption does not hold any more. Finally, $\mathbb{V}[Y_i(g)] \neq \mathbb{V}\left[\bar{Y}_i(g)\right]$, even in the specific case where $\forall i, j, g, g'$, $\mathrm{Cov}[\mu_i(X_g), \mu_j(X_{g'})] = 0$. 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 $X_{g'}$, it is easy to define a mixture r.v. $X_{g \cup g'}$. From $\mathbf{Y}(X_g)$ and $\mathbf{Y}(X_{g'})$, one can thus build $\mathbf{Y}(X_{g \cup g'})$. 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].*

*Mixture approach has no smoothing effect since it includes by construction the population diversity, resulting in stabilized correlations between output variables. In other words, there is no difference between output variables for a grain seen as an aggregation of 2 grains and the same set of points seen as a single grain.*

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 1.

**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 :*

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

- *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 :*

$$\begin{cases} X_g &= BX_{g_0} + (1-B)X_{g\setminus g_0} \\ X_{g'}^{\perp} &= B'X_{g_0}^{\perp} + (1-B')X_{g'\setminus g_0} \end{cases} \tag{6}$$

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

$$\Delta := \mathrm{Cov}\left[Y_i(X_g), Y_j(X_{g'}^{\perp})\right] - \mathrm{Cov}\left[Y_i(X_g), Y_j(X_{g'})\right]. \tag{7}$$

*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*

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

*one can show that:*

$$|\Delta| \leq \mathbb{P}\left[B = B' = 1\right]\mathbb{P}\left[X_{g_0} \neq X_{g_0}^{\perp}\right]\rho_{\max}\ . \tag{8}$$

*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. In other cases, one may use results in Proposition 1.*

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

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

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

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

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 $\rho_{\max}$ , hence the result. □

# 3 Prediction

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

For $(i_1, \ldots, i_n) \in [\![1,p]\!]^n$ and $g_1, \ldots, g_n \in \mathcal{G}$ we know $n$ random variables :

$$\underline{\mathbf{Y}} = (Y^1, \ldots, Y^n)^{\top} \quad \text{with} \quad Y^j = Y_{i_j}(g_j) \text{ for } j \in [\![1,n]\!]$$

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

$$\underline{\mathbf{Y}} = (Y_1(X_{g_1}), \ldots, Y_p(X_{g_1}), \ldots, Y_1(X_{g_j}), \ldots, Y_p(X_{g_j}), \ldots, Y_1(X_{g_k}), \ldots, Y_p(X_{g_k}))^{\top}. \tag{9}$$

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 $\underline{\mathbf{Y}}$ as given in Equation (9). 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 $\mathbf{Y}(g) = (Y_1(g), \ldots, Y_p(g))^{\top}$, the problem amounts to predict $Y_i(g)$ for all $i \in [\![1,p]\!]$.

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

We denote :

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

In the following, we assume that $\mathbf{K}$ is invertible. In particular, the problem of repeated observations is not treated here. Note that if the expectations of $Y_i(x)$ and covariances between $Y_i(x)$ and $Y_j(x')$ are known for all $i, j \in [\![1,p]\!]$, $x, x' \in \chi$, as in usual Kriging assumptions, then by Proposition 2, $\underline{\boldsymbol{\mu}}$, $\mathbf{K}$ and $\mathbf{h}_i(g)$ can be computed.

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

$$M_i(g) = \sum_{j=1}^{n} \alpha_i^j(g) Y^j = \boldsymbol{\alpha}_i(g)^\top \underline{\mathbf{Y}}. \tag{10}$$

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

$$\boldsymbol{\alpha}_i(g) \in \arg\min_{\boldsymbol{\alpha} \in \mathbb{R}^n} \mathbb{E}\left[\left(Y_i(g) - \boldsymbol{\alpha}^\top \underline{\mathbf{Y}}\right)^2\right] \tag{11}$$

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

$$\left\{\begin{array}{lcl}
\epsilon_i(g) & := & Y_i(g) - M_i(g) \\
c_{i,j}(g,g') & := & \mathbb{E}\left[\epsilon_i(g)\,\epsilon_j(g')\right] \\
v_i(g) & := & c_{i,i}(g,g)
\end{array}\right. \tag{12}$$

**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:*

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

*And similarly, we denote:*

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

*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:*

$$\forall i, j \in [\![1,p]\!],\ \forall g, g' \in \chi,\ F_i(g) \parallel F_j(g')$$

*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 3.*

The following proposition gives an optimal predictor that can be computed under the minimal assumptions of Proposition 2 : given the first two moments of random variables $\{X_g \;:\; g \in \mathcal{G}\}$, all components of $\underline{\boldsymbol{\mu}}$, $\mathbf{K}$ and $\mathbf{h}_i(x)$ can be computed.

**Proposition 3** (Mixture Kriging prediction)**.** *Given a set of observations $\underline{\mathbf{Y}}$, for any $g \subset \chi$, and in particular for a single point $g = \{x\}$, for any $i \in [\![1,p]\!]$, the weights $\boldsymbol{\alpha}_i(g)$ yielding the best linear unbiased predictor (BLUP) of $Y_i(g)$ and the associated cross errors are as follows:*

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

$$\left\{\begin{array}{lcl}
\boldsymbol{\alpha}_i(g) & = & \mathbf{K}^{-1}\mathbf{h}_i(g) \\
c_{i,j}(g,g') & = & k_{i,j}(g,g') - \mathbf{h}_i(g)^\top \mathbf{K}^{-1}\mathbf{h}_j(g')
\end{array}\right.$$

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

$$\left\{\begin{array}{lcl}
\boldsymbol{\alpha}_i(g) & = & \mathbf{K}^{-1}\left(\mathbf{h}_i(g) + \lambda_i(g)\underline{\boldsymbol{\mu}}\right) \quad \textit{where} \quad \lambda_i(g) = \frac{\mu_i(g) - \underline{\boldsymbol{\mu}}^\top \mathbf{K}^{-1}\mathbf{h}_i(g)}{\underline{\boldsymbol{\mu}}^\top \mathbf{K}^{-1}\underline{\boldsymbol{\mu}}} \\
c_{i,j}(g,g') & = & k_{i,j}(g,g') - \mathbf{h}_i(g)^\top \mathbf{K}^{-1}\mathbf{h}_j(g') + \lambda_i(g)\lambda_j(g)\underline{\boldsymbol{\mu}}^\top \mathbf{K}^{-1}\underline{\boldsymbol{\mu}}
\end{array}\right.$$

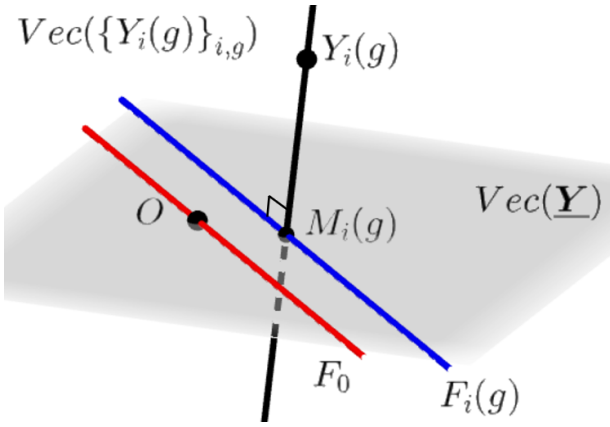


Figure 3: *Geometrical interpretation of the prediction process.*

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

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

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

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

By differentiation over each component of $\boldsymbol{\alpha}$,

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

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

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

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

(ii) If $\underline{\boldsymbol{\mu}} \neq (0, \ldots, 0)^\top$ 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 :

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

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:

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

Due to unbiasedness condition, it means that:

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

Which rewrites:

$$c_{i,j}(g,g') = k_{i,j}(g,g') - \boldsymbol{\alpha}_j(g')^\top \mathbf{h}_i(g) - \boldsymbol{\alpha}_i(g)^\top \mathbf{h}_j(g') + \boldsymbol{\alpha}_i(g)^\top \mathbf{K} \boldsymbol{\alpha}_j(g'). \tag{13}$$

Note that equation (13) 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. □

**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*

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

*we can write :*

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

*therefore yielding an unbiased estimator of the unknown common expectation.*
*To be compared with usual sample mean for independent observations:*

$$\overline{\underline{\mathbf{Y}}} = \frac{\mathbf{1}_n{}^\top \underline{\mathbf{Y}}}{\mathbf{1}_n{}^\top \mathbf{1}_n}.$$

**Remark 5** (Uncertainty on outputs). *In practice, each observation of $Y_g = Y(X_g)$ may consist in a mean and a variance, e.g. the mean salary over a state, the variance of this salary over the state. The latter variance may come:*

1. *from the variation of $X_g$ over the considered area $g$, propagated to $k_{i,i}(g,g)$ via Proposition 2,*
2. *or from a nugget effect $\epsilon_x$ added on pointwise locations $k_{i,i}(x,x) + \epsilon_x$, $x \in g$ and thus propagated to $k_{i,i}(g,g)$ via Proposition 2,*
3. *or even from a nugget effect $\epsilon_g$ directly added to $k_{i,i}(g,g) + \epsilon_g$.*

*The modelling of this variation source depends on the application, but in all cases the results presented in Proposition 3 are directly applicable to the modelled covariances, as long as the resulting covariance matrix $\mathbf{K}$ is definite positive (and it must be, as it models a covariance between properly defined random variables). One just has to check that all combined sources of variations generate the desired variance over the grain.*

**Remark 6** (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*

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

*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. This last condition holds in particular when each grain* $g$ *is reduced to one point. In this setting, under assumption* (A)*, one can show that cross errors for both Simple Mixture Kriging and Ordinary Mixture Kriging are*

$$c_{i,j}(g,g') = \mathbb{E}\left[\mathrm{Cov}\left[Y_i(g), Y_j(g')|\underline{\mathbf{Y}}\right]\right]. \tag{14}$$

*If* $\mathrm{Cov}\left[Y_i(g), Y_j(g')|\underline{\mathbf{Y}}\right]$ *does not depend on* $\underline{\mathbf{Y}}$*, as it is the case for conditional Gaussian vectors, Equation* (14) *simplifies:* $\mathbb{E}\left[\mathrm{Cov}\left[Y_i(g), Y_j(g')|\underline{\mathbf{Y}}\right]\right] = \mathrm{Cov}\left[Y_i(g), Y_j(g')|\underline{\mathbf{Y}}\right]$.

*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 $\beta$ 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 $v_i(g) := \mathbb{E}\left[(\epsilon_i(g))^2\right]$. We have:

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

  The minimum value of this polynomial expression is reached for:

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

  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 $\mathbb{E}\left[\epsilon_i(g)\right] = 0$ and $\mathbb{E}\left[\delta\right] = 0$:

  $$\forall \delta \in F_0,\ \forall i \in [\![1,p]\!],\ \forall g \in \chi,\quad \mathbb{E}\left[\epsilon_i(g)\,\delta\right] = \mathrm{Cov}\left[\epsilon_i(g), \delta\right] = 0. \tag{15}$$

  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 $\delta$ and $\delta'$ be any two vectors of $F_0$. As a consequence of the previous result in Equation (15), we have:

  $$\mathrm{Cov}\left[\epsilon_i(g) + \delta, \epsilon_j(g') + \delta'\right] = c_{i,j}(g,g') + 0 + 0 + \mathrm{Cov}\left[\delta,\, \delta'\right] \tag{16}$$

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

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

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

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

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

  $$\mathrm{Cov}\left[\epsilon_i(g) + \delta, \epsilon_j(g') + \delta'\right] = \mathbb{E}\left[\mathrm{Cov}\left[Y_i(g), Y_j(g') \mid \underline{\mathbf{Y}}\right]\right] + \mathrm{Cov}\left[\delta, \delta'\right] \tag{17}$$

  Identifying the equations (16) and (17), we get the expected result.

□

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

# 4 Numerical illustrations

The numerical application provided here shows that the results presented in this paper can be applied to solve some uncertainty propagation problems. We use here of some test functions rather on a real dataset, in order to make the application easier to understand. The illustration is presented here in an industrial context, but can be easily adapted to other contexts: geographical, ecological, etc.

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 region $g_x \subset \chi$ around the design $x \in \chi$.

As a consequence, when testing different possible designs $x_1, \ldots, x_k$, the industry observes $\mathbf{Y}(g_1)$, $\mathbf{Y}(g_k)$. 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 $x_1, \ldots, x_k$. Starting from given covariances between points, we have built the grain covariances of Proposition 2. In a second step, we have built the predictor presented in this paper in Proposition 3. The results appear in Figures 4, 5 and 6.The setting details are given in figures captions.

In Figure 4, one sees that when grains are singletons, classical Kriging results are retrieved. When grains are not restricted to singletons, 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 clearly depends on the size of the grain. Reducing a grain to its centroid ends up in a very different model (see e.g. area-to-area Kriging in Goovaerts [2006].

In Figure 6, we show the interest of being able to integrate aggregated data with pointwise observations. We suppose that we know the expectation and variance of $y$ on some given grains. We add observations to the learning set on the grains with value the expectation of $y$. We add the variances to the diagonal terms of the covariance matrix $\mathbf{K}$ associated with the grains. We observe a great reduction of uncertainty in the second case. However we also observe an adverse edge effect of this approach as the prediction is changed on the left side of the right graphic (see Remark 5).

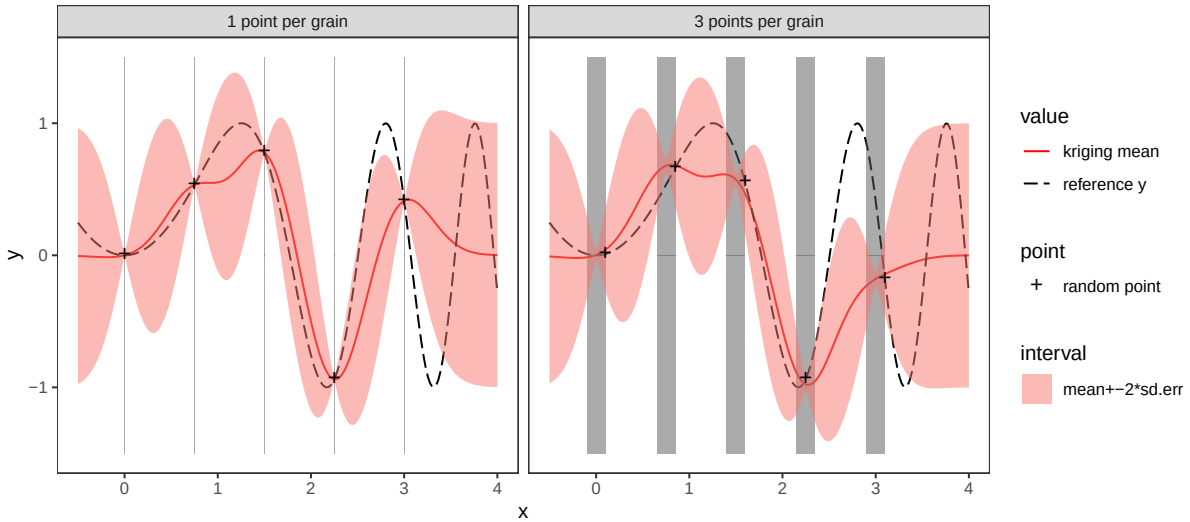


Figure 4: *Comparison of classical simple Kriging and mixture simple Kriging.* **Left:** Reference function $y$ (solid line) is observed at 5 points without uncertainty on the value of $x$. We find the usual results on simple Kriging, in particular predictions (dashed line) are interpolating. **Right:** 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, $r = 0.1$.

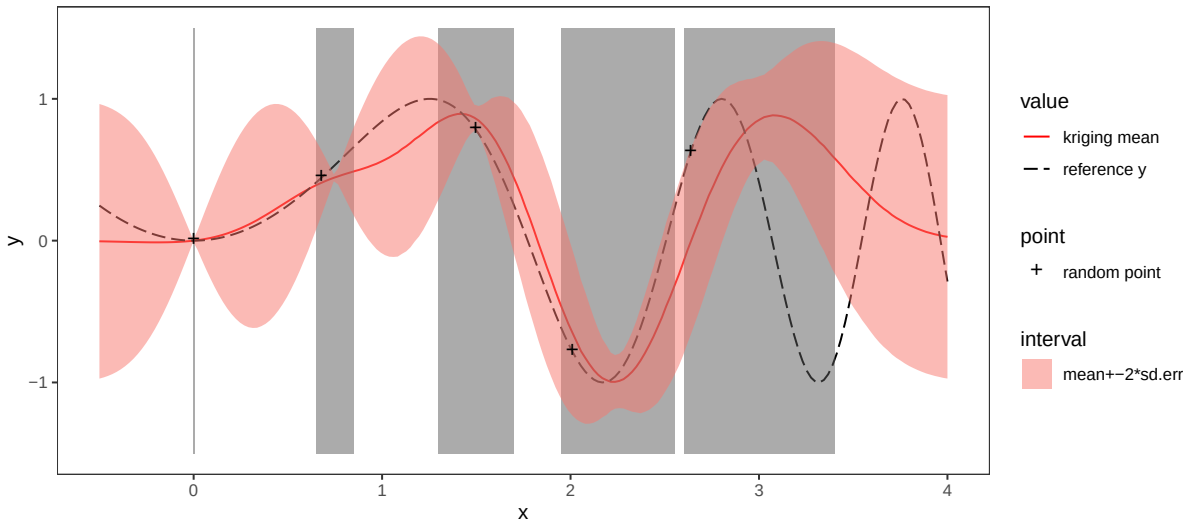


Figure 5: *Varying grain sizes and effect on uncertainty.* Reference function $y$ (solid line) is observed at 5 grains of different sizes. We observe that predictions uncertainty is growing when the grain size is growing. **Details:** In this example, $y_{\text{reference}}(x) = sin(x^2)$, covariance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

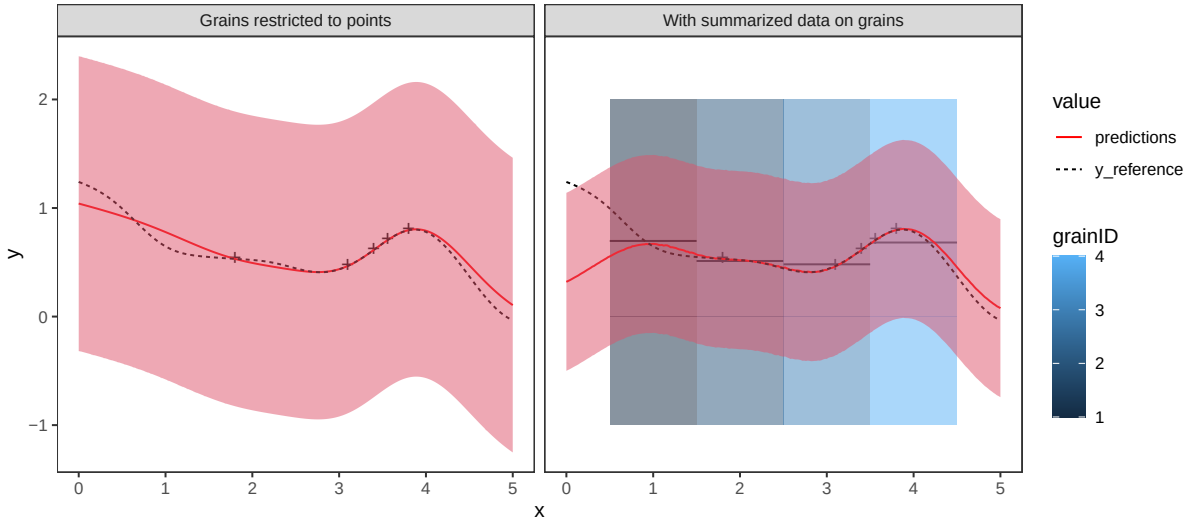


Figure 6: *Effect of knowing aggregated information on a grain.* **Left:** Ordinary Kriging in its classical form, with grains restricted to points. **Right:** In addition to the observations on points, we assume that we know expectation and variance of $y$ over 4 grains. We integrate this data both 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.


# 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.


# 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.

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.

Osvaldo 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.

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

*Proof of Proposition 1.*

- 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 $g_+ \in \underset{g \in \mathcal{G} \cup \{h\}}{\cup} g$. Moreover:

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

  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:

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

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

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

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

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

- 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:

  $$\text{(A)} \;\; \exists h \in \mathcal{H} \; : \; g \subset h \cap g_0 \quad \text{or} \quad \text{(B)} \;\; \exists h \in \mathcal{H} \; : \; g \subset h \setminus g_0 \quad \text{or} \quad \text{(C)} \;\; g \subset g_0 \setminus \underset{h \in \mathcal{H}}{\cup} h$$

  $$\text{and} \quad \text{(a)} \;\; \exists h' \in \mathcal{H} \; : \; g_+ = h' \cap g_0 \quad \text{or} \quad \text{(b)} \;\; \exists h' \in \mathcal{H} \; : \; g_+ = h' \setminus g_0 \quad \text{or} \quad \text{(c)} \;\; g_+ = g_0 \setminus \underset{h \in \mathcal{H}}{\cup} h$$

  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 $\mathcal{G} \leq \mathcal{H} \oplus \{g_0\}$.

□

# Contents


**1 Introduction** **1**

**2 Spatial model** **4**
2.1 Inputs . . . . . 4
2.2 Outputs . . . . . 5

**3 Prediction** **8**

**4 Numerical illustrations** **13**

**A Appendix** **17**