

1 Predicting missing Energy Performance Certificates:
2 Spatial interpolation of mixture distributions

3 Marc Grossouvre^{a,b}, Didier Rullière^c, Jonathan Villot^d

^a*Mines Saint-Etienne, CNRS, UMR 6158 LIMOS, Institut Henri Fayol, Departement
GMI, Espace Fauriel, 29 rue Ponchardier, Saint-Etienne, 42023, France;*

^b*U.R.B.S. SAS, Bâtiment des Hautes Technologie, 20 Rue Professeur Benoit
LAURAS, Saint-Etienne, 42000, France*

^c*Mines Saint-Etienne, Univ Clermont Auvergne, CNRS, UMR 6158 LIMOS, Institut
Henri Fayol, Espace Fauriel, 29 rue Ponchardier, Saint-Etienne, 42023, France*

^d*Mines Saint-Etienne, Univ Lyon, CNRS, Univ Jean Monnet, Univ Lumiere Lyon 2,
Univ Lyon 3 Jean Moulin, ENS Lyon, ENTPE, ENSA Lyon, UMR 5600 EVS, Institut
Henri Fayol, Saint-Etienne, 42023, France*

4 **Abstract**

Mass renovation goals aimed at energy savings on a national scale require a significant level of public financial commitment. To identify target buildings, decision-makers need a thorough understanding of energy performance. Energy Performance Certificates (EPC) provide information about areas of space, such as land plots or a building's footprint, without specifying exact locations. They cover only a fraction of dwellings. This paper demonstrates that learning from observed EPCs to predict missing ones at the building level can be viewed as a spatial interpolation problem with uncertainty both on input and output variables. The Kriging methodology is applied to random fields observed at random locations to determine the Best Linear Unbiased Predictor (BLUP). Although the Gaussian setting is lost, conditional moments can still be derived. Covariates are admissible, even with missing observations. We present applications using both simulated and real data, with a specific case study of a city in France serving as an example.

5 *Keywords:* climate governance, energy efficiency, multi-scale processes,
6 areal data, change of support

Email addresses: marcgrossouvre@urbs.fr (Marc Grossouvre),
drulliere@emse.fr (Didier Rullière), jonathan.villot@emse.fr (Jonathan Villot)

7 **1. Introduction**

8 *1.1. Classifying the EPC prediction problem in research*

9 Energy Performance Certificates are delivered in many countries around
10 the world to assess the energy efficiency of buildings. Various approaches can
11 be found in the European Union, Turkey, the UK, the USA... An Energy
12 Performance Certificate (EPC) is defined in France as an energy consumption
13 associated with a qualitative labelling letter ranging from A to G as shown in
14 Figure 1. Energy consumptions associated with dwellings, identified by their
15 addresses, are inventoried in a database released in open access and mapped
16 in Figure 2. A second database matches each address with a land plot.
17 Finally, a third database gives the living area of every dwelling, be it house
18 or apartment, together with the land plot where they are located, and a few
19 other technical specifications. However, the exact location of these dwellings
20 on each land plot is not certain. From these datasets, decision-makers such as
21 municipalities, would like to infer the EPC (energy consumption and label) of
22 buildings that have not been observed in order to identify targets for energy
23 retrofit incentives. This problem is referred to as the EPC prediction problem
24 in the present paper.

25 In the literature, this problem can be approached from an engineering
26 perspective, from a data management perspective or from a geostatistics
27 point of view.

28 From an engineering perspective, heat engineers have physical models
29 that compute an energy balance in order to find a given building's energy
30 consumption. These models require a large number of technical features
31 and may be used to design a refurbishment (improvement) strategy (Baker
32 et al., 2021). To work at a larger scale, heat engineers define typologies of
33 buildings, compute a distribution of these types on a given territory, and
34 therefore infer a distribution of EPC labels. This approach has proven to
35 be efficient (Ballarini et al., 2017). However, the lack of knowledge about
36 the detailed technical features of each building is a strong limitation for a
37 prediction at the building level. Some feature reduction efforts have been
38 made (Ali et al., 2020), but the remaining features are still problematic to
39 infer and require extra efforts (Schetelat et al., 2020). The present work
40 considers an alternative approach wherein detailed technical knowledge of
41 each building is relinquished, and instead leverages the geolocated nature of
42 EPC information.

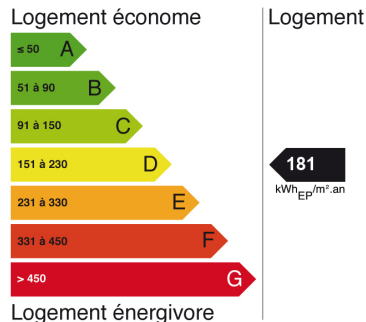


Figure 1: Prescribed vignette appearing on the French energy performance certificate up to 2021. Label A refers to energy-efficient dwellings, and label G refers to energy-intensive dwellings.

43 From a data management perspective, the EPC prediction problem re-
 44 quires a process to combine datasets from multiple sources available at mul-
 45 tiple scales, which is known as data fusion (Smith et al., 2008). These types
 46 of problems are becoming increasingly complex due to the growing amount
 47 of data available, whether it be ecological, social, or institutional. These
 48 datasets relate to space units of varying shapes, dimensions, and cardinal-
 49 ity. And in some cases, it may be difficult to determine the exact position
 50 of an observed object. This is the case with buildings, since many govern-
 51 ments lack a detailed map of the building stock in their country. Property
 52 tax is typically based on intrinsic factors such as surface area and number
 53 of bedrooms, but not extrinsic factors such as the floor number or window
 54 orientation (see Table 1). As a result of this uncertainty, large-scale studies
 55 on housing stock have to rely on an abstract concept of dwelling. This idea of
 56 dwelling can refer to a house or an apartment; it is not clearly delimited but
 57 it is described by a set of features such as an area or a number of bedrooms.
 58 These features are gathered in a table with one dwelling per row, meaning
 59 that the dwelling is the smallest unit of information.

dwelling ID	address	area (m^2)	bedrooms	...	land plot ID
024830065432	161 rue du Chateau 02089 BILLY	83	3	...	024830000C0057

Table 1: Structure of the dwellings table from the French Ministry of Finance. The actual table comprises 118 features. Geographic position is identified by a land plot ID.

60 Similarly, the smallest unit of information for a table with one EPC per
 61 row is a part of a building. It is not clearly defined as an object in a 3-

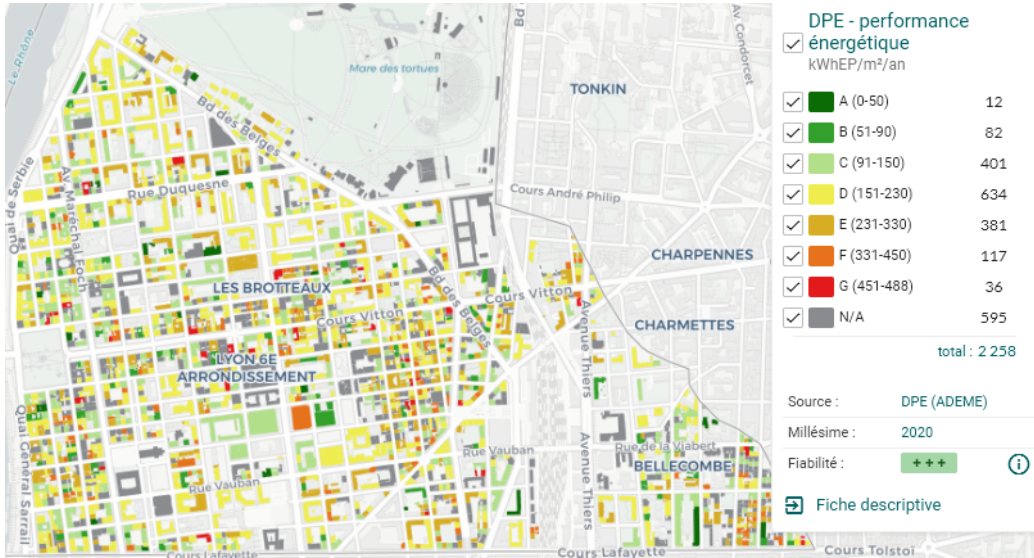


Figure 2: Map of French inventoried EPCs over a neighborhood of Lyon city. This image is a screen capture of the French National Observatory of Buildings (Observatoire National des Bâtiments - ONB), released with the consent of the rights holders U.R.B.S. SAS.

address	area (m^2)	walls	...	energy consumption	EPC
161 rue du Chateau 02089 BILLY	83	bricks	...	210	D

Table 2: Structure of the observed EPCs table. Geographic positions are indicated by addresses.

62 dimensional space, but it has features that describe it (see Table 2). And
 63 to predict the EPC of buildings, one also has to define buildings. In the
 64 same way, data fusion requires defining the smallest units of information,
 65 also known as granules for each dataset: “Informally, a granule of a variable
 66 X is a clump of values of X that are drawn together by indistinguishability,
 67 equivalence, similarity, proximity, or functionality. For example, an interval
 68 is a granule.” Zadeh (2005). The field of study that focuses on representing,
 69 constructing, and processing these information granules is called Granular
 70 Computing (Pedrycz, 2013). Assuming that an appropriate data fusion pro-
 71 cess is implemented, dwellings, EPC observations, and complete buildings
 72 are represented in the same data model. It remains to define a relevant
 73 predictive model. Granular computing is multidisciplinary, but since we are
 74 dealing with geo-localized information, the natural field of research is geo-
 75 statistics, which has been defined as “dealing with spatial processes indexed

76 over continuous space” (Cressie, 1993, p7).

77 From a geostatistics perspective, the irreducible uncertainty about gran-
78 ules’ positions (dwellings, buildings, etc.) in their underlying space restricts
79 the use of traditional spatial interpolation models such as Kriging as well as
80 more recent models such as those proposed by Roksvåg et al. (2021), although
81 the latter efficiently combines point and areal observations. This work aims
82 to overcome the latter limitation and develop a comprehensive framework
83 capable of handling data with uncertainty about the position of observed
84 objects while still allowing for the definition of an optimal linear predictor
85 for spatial interpolation of EPC values. As is first presented below, the liter-
86 ature shows that the problems to solve have already been identified and that
87 several solutions have been proposed with their benefits and shortcomings.

88 *1.2. The limits of systematic averaging for spatial interpolation*

89 Gaussian Process Regression (Williams and Rasmussen, 1996), also known
90 as Kriging, is one of the major spatial interpolation approaches (Comber and
91 Zeng, 2019). Kriging theory relies on the assumption that points close to each
92 other are more likely to have similar features. It achieves the Best Linear
93 Unbiased Predictor (BLUP) in the least squares sense for point spatial inter-
94 polation. However, the EPC prediction problem deals with observations that
95 are not point observations but areal observations. Areal interpolation, as de-
96 fined by Lam (1983), involves “the transformation of data from one set of
97 boundaries to another”. Lam also used the terms source zone and target zone.
98 For the EPC prediction problem, source zones are dwellings and buildings’
99 parts that are observed, while target zones are whole buildings, including
100 those for which no part has been observed. Spatial or areal interpolation re-
101 search is based on the assumption that granules close to each other are more
102 likely to have similar features. This is reasonably understandable for temper-
103 atures that are continuously defined over space, but it may be more challeng-
104 ing to observe and model when dealing with areal data where granules can
105 be of various sizes and shapes, sometimes uncertainly defined. Gotway and
106 Young (2002) highlighted the terms used to describe areal interpolation and
107 its challenges; this terminology includes block Kriging, multi-scale and multi-
108 resolution modelling, the ecological inference problem, the modifiable areal
109 unit problem (MAUP), the scaling problem, the change of support problem,
110 and the reduction of variance problem. Below are the aspects of this work
111 that are more relevant for solving the EPC prediction problem.

112 Block Kriging is a derivative of Kriging designed for handling areal data.
113 It distinguishes point-to-area, area-to-point, and area-to-area predictions. It
114 assumes that a feature at block (granule) level is the average of the block’s
115 point features. Point-to-area prediction produces an estimate “identical to
116 that obtained by averaging the point estimates produced by [Kriging]” (Isaaks
117 and Srivastava, 1989; Cressie, 1993). Kyriakidis (2004) described a complete
118 Kriging model for area-to-point prediction, proved that it is an optimal pre-
119 dictor, and sketched area-to-area prediction. Goovaerts (2008) studied in
120 depth the problem of estimating the variogram, that is to say, measuring
121 the similarity between 2 points at different distances, for block Kriging. He
122 showed that averaging reduces the sill of the variogram and tried to tackle
123 this bias. Moreover, while point estimates obtained by Kriging are optimal,
124 area-to-area Kriging may not be the optimal predictor for the average value
125 over the block.

126 A known issue resulting from systematic averaging in areal Kriging models
127 arises in scenarios such as analysing crop yields, where the set of agricultural
128 fields to aggregate for a certain type of crop varies from year to year. It states
129 that correlations between features at areal level are heavily dependent on the
130 aggregation process, making it difficult to compare correlations between dif-
131 ferent years. This is the Modifiable Areal Unit Problem (MAUP) for which a
132 measuring approach has been recently proposed (Briz-Redon, 2022). While
133 the MAUP refers to the correlation between aggregated features, the ecolog-
134 ical inference problem is a result of the correlations at the individual level
135 being different from the correlations of the averaged features at the ecological
136 (group) level. A lack of information about the individuals’ positions leads
137 to a bias when the averaged information about individuals distributed into
138 areal units is cross-classified by other individual (point-level) variables (sex,
139 race). According to Gotway and Young (2002), “The smoothing effect that
140 results from averaging is the underlying cause of both the scale problem in the
141 MAUP and aggregation bias in ecological studies.” Apart from correlations,
142 the variance itself is affected by systematic averaging. Indeed, the average
143 of identical random variables has a smaller variance than the variance of the
144 individuals themselves. The specific issue of variance reduction at the block
145 level was partially addressed in Li et al. (2009) using rectangular blocks at
146 multiple scales.

147 Despite its limitations, the averaging method has proven to be effective for
148 interpolating areal data. For example, Poggio and Gimona (2015) downscaled
149 climate models and predicted soil wetness using Kriging on the residuals

150 of a generalized additive model (Wood, 2017). Area-to-point Kriging, also
151 called disaggregation, has also been implemented by Kerry et al. (2013);
152 Truong and Heuvelink (2013); Yoo and Kyriakidis (2006). Additionally, area-
153 to-area Kriging (block Kriging) has been used effectively by Zhang et al.
154 (2018) and has been apply to downscaling by Jin et al. (2018) as well as
155 Pereira et al. (2018). The satellite imaging field has also notably benefited
156 from this framework, as in the pan-sharpening process, which is “a technique
157 to combine the fine spatial resolution panchromatic (PAN) band with the
158 coarse spatial resolution multispectral bands of the same satellite to create
159 a fine spatial resolution multispectral image” Wang et al. (2016). In this
160 process, points are weighted according to their distance from the centroid of
161 the satellite pixel when computing the average value.

162 Both the MAUP and the ecological inference problem belong to a fam-
163 ily of problems related to the combination of different types of granules in
164 the same model, e.g. observing dwellings and predicting buildings. These
165 problems are gathered in the change of support problems family. Another
166 particular change of support problem known as spatial misalignment arises
167 when a given feature is observed at multiple scales, including point level.
168 Systematic averaging makes points and areas different objects with differ-
169 ent different correlation structures and therefore different predictors. The
170 classification of problems such as “area-to-point” or “area-to-area” reflects
171 this categorization. To address spatial misalignment, a Bayesian framework
172 that can be iterated both with point observations and block observations has
173 been proposed by Moraga et al. (2017). However, this model is still based
174 on averaging at areal level for features that are continuously defined over the
175 territory. Like other models derived from Kriging, it considers blocks to be
176 connected surface areas in \mathbb{R}^2 that need to be discretized (Goovaerts, 2008),
177 which can distort reality for features that are not continuously defined over
178 the space. Such is the case of populations of individuals that are discrete
179 points heterogeneously located within a block, such as a county or census
180 tract.

181 *1.3. Beyond systematic averaging*

182 A way to try and overcome change of support problems is to define a new
183 data model for which features at areal level do not require systematic aver-
184 aging. In this regard, Godoy et al. (2022) defined a Gaussian random field
185 on the class \mathcal{B}_D of closed subsets of a certain domain $D \in \mathbb{R}^n$. Distances

186 between elements of \mathcal{B}_D are measured with the Hausdorff distance, and the
187 correlation structure between features is based on this distance together with
188 a Matérn kernel. Eventually, a Bayesian framework is used to fit the model
189 with respiratory cancer data, yielding encouraging results. This model seems
190 very general and will probably find other fields of application. However, it is
191 not interpretable in the sense that there is no obvious link between the fea-
192 ture at the areal level and the feature at the point level, therefore eluding the
193 question of consistency. In other words, it is not known whether the aggrega-
194 tion of cancer incidence predictions at a small scale would give the prediction
195 of cancer incidence at a larger scale. Beside this limitation, the Hausdorff-
196 Gaussian process does not solve the problem of position uncertainty that is
197 found in the EPC prediction problem.

198 In this paper, a new model is proposed where learning and prediction can
199 be made from both aggregated and point support data. An object category
200 called grain is introduced to express this new approach, consistent with re-
201 search realities where it may be desirable to complete large aggregated open
202 datasets with local observations and predict at various scales. Grains con-
203 taining a continuous or discrete set of points are treated identically. As is
204 detailed above, a weighted average is the standard aggregation approach.
205 In this respect, the MAUP is related to determining a covariance model for
206 points from which the covariance between blocks and the covariance between
207 points and blocks are derived. Weights for averaging are assumed to be fully
208 determined for a given block; they are not regarded as a probability distribu-
209 tion for a block, thereby ignoring some related statistics and other potential
210 sources of stochastic dependence between blocks. The present paper pro-
211 poses a method of incorporating a mixture distribution to address this issue.
212 Kriging has already been developed for features that are mixtures at the
213 point level (Lin et al., 2010), but Lin et al. make no assumption about the
214 distribution of features at the areal level. Instead, we assume the aggregation
215 of information at the areal level to be a mixture. Averaging a large number
216 of random variables results in a variance reduction, whereas mixing a large
217 number of random variables does not tend to reduce the variance. We will
218 show that this approach effectively manages position uncertainty. However,
219 one drawback is that mixtures of Gaussian random variables are generally
220 not Gaussian, which means that the usual Gaussian process interpretations
221 and conditioning will no longer hold.

222 The present study proposes a new model for processing granular data,

223 as detailed in Section 2. In Subsection 2.1, a suitable data model is estab-
224 lished, while in Subsection 2.2, we define the feature variables' means and
225 covariances. Moreover, a Best Linear Unbiased Predictor is derived in Sub-
226 section 2.3. We illustrate the model with examples in Section 3, starting
227 with simulated rounded positions in Subsection 3.1, followed by simulated
228 areal data with varying area sizes in Subsection 3.2. Subsection 3.3 focuses
229 on presenting the EPC prediction problem with real data. Finally, in Section
230 4, we discuss the pros and cons of the new model.

231 2. Prediction model

232 This work is motivated by the will to handle data that is released in open
233 format by public or private institutions. The goal is to use institutional data,
234 such as the distribution of salaries at the municipality level, to estimate the
235 distribution of salaries at a smaller scale, such as a district in a city, while
236 also including known salaries at specific locations. To achieve this, we pro-
237 pose here a general Kriging approach that extends the traditional Simple or
238 Ordinary Kriging and coKriging techniques. The model will explain some
239 variables (such as the energy consumption, the salary, etc.) using some ex-
240 planatory variables (such as the location, the construction year, etc.). The
241 former will be referred to as output variables and the latter as input vari-
242 ables. Let us consider a space (input space, sometimes known as study space)
243 over which is defined a field of multidimensional random variables (output
244 variables, features of interest) such as sociological variables, assumed to be
245 defined and potentially observed for both points in the input space and for
246 geographic areas, such as cities, regions, or countries. These areas are re-
247 ferred to as “grains”. The model predicts output variables at unobserved
248 points or grains, based on the assumption that the dependence between out-
249 puts depends on the relative positions of the inputs. No assumption is made
250 regarding the shape of the grains, which can even overlap partially or com-
251 pletely.

252 2.1. Data model

253 Let us define the structure of the input space.

254 **Definition 1** (Inputs). *Let d be a positive integer. A territory and grains*
255 *inside this territory are defined as follows:*

- 256 • A **territory** is a subset χ of \mathbb{R}^d ;

- 257 • A *point* is any element $x \in \chi$;
- 258 • A *grain* is any non-empty subset $g \subseteq \chi$.

259 It is common in some application fields to use a different terminology
 260 to talk about grains: blocks, pixels, or areas for instance. In the above
 261 definition, there is no constraint on grains, contrary to pixels that are usually
 262 forming a regular grid known as a raster. A set of grains does not have to
 263 cover the whole territory, and its elements might overlap. Moreover, a grain
 264 is not necessarily a connected set, contrary to blocks. And an area is usually
 265 seen as associated with a surface area (a set of strictly positive measure)
 266 whereas a grain may be a finite set of points.

267 For instance, suppose that the points are represented as pairs of latitude
 268 and longitude coordinates in an appropriate coordinate reference system. In
 269 this case, χ could be defined as the set of all latitude-longitude pairs that
 270 fall within a specific country, yielding $d = 2$ and $\chi \subset \mathbb{R}^2$. A grain may
 271 correspond, for example, to a specific city, to a specific land plot, or to a
 272 specific building's footprint. Previous Kriging models refer to blocks or areas
 273 for sets of points that are disjoint (see, for instance, Kyriakidis, 2004).

274 When dealing with geographic data, a set of grains is usually the minimum
 275 scale at which information is available; that is to say, the data granularity.
 276 For instance, it may be the set of land plots, the set of cities, the set of
 277 buildings' footprints, etc. However, considered grains may have non-empty
 278 intersections and may come from different datasets, at different scales, such
 279 as land plots and census tracts. Definition 1 is general enough to include
 280 such sets of grains. Data that describe population or buildings are not con-
 281 tinuously defined over a territory, as opposed to temperature or pollutant
 282 concentration. Census data are anonymized at the census tract level before
 283 being released. For instance, in a census table describing dwellings, a row
 284 describes a dwelling that exists on a certain census tract, but we don't know
 285 exactly where it is on this tract. Then dwellings' surface area is neither con-
 286 tinuous nor clearly geo-localized. Definition 2 below unifies output features
 287 that are continuously defined over a territory and output features that are
 288 not.

289 An originality of this work is to consider a set of random locations that
 290 model the uncertainty of explanatory variables over each considered grain.
 291 Let $\{X_g, g \in \mathcal{G}\}$ be a given sequence of random variables that are random
 292 locations, or more generally, random explanatory variables. Their joint distri-

293 bution is assumed to be known. As an example, for non-overlapping grains,
 294 one can choose a sequence of independent uniform random variables over
 295 each grain, but any other joint distribution, possibly dependent, can be cho-
 296 sen. Definition 2 associates output variables with these random explanatory
 297 variables.

298 **Definition 2** (Outputs). *Let \mathcal{G} be a set of grains. We assume that for each*
 299 *grain $g \in \mathcal{G}$, there is a random variable X_g with values the points of g . Output*
 300 *variables are defined over points and grains of \mathcal{G} as follows:*

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

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

- 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 $X_g \in g$:

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

301 *Defined accordingly, $\mathbf{Y}(g)$ is a mixture distribution.*

302 *For a given set of grains \mathcal{G} , the set of random variables $\{X_g : g \in \mathcal{G}\}$, is*
 303 *assumed to be defined and known, and the dependence structure between*
 304 *those random variables is supposed to be known. Furthermore, these*
 305 *random variables are assumed to be independent from the random field*
 306 \mathbf{Y} .

307 **Example 1.** *The importance of X_g should be stressed here. For instance,*
 308 *if one studies the distribution of capital owned by citizens of a given mu-*
 309 *nicipality, $P(X_g = x)$ gives the probability of a citizen x to be observed.*
 310 *$P(Y(X_g) = y)$ is the probability to observe y when a citizen picked randomly*
 311 *according to X_g unveils his capital:*

$$P(Y(g) = A) = P(Y(X_g) = A) = \sum_{x \in g} P(X_g = x)P(Y(x) = A) .$$

312 *It is clear that individuals are not distributed regularly (along a grid for in-*
 313 *stance) in the grain. However, in this example, it makes sense to consider*
 314 *that $\forall x \in g, P(X_g = x) = 1/[g]$ where $[g]$ is the cardinality of g . This means*
 315 *that the contribution of all citizens are equally valued in $Y(g)$.*

Let us now suppose that the outputs are partially known on a set of grains. For $(i_1, \dots, i_n) \in \{1, \dots, p\}^n$ and $g_1, \dots, g_n \in \mathcal{G}$ the following n random variables are known:

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

316 As an example, if ℓ observations of the whole random vector $\mathbf{Y}(g_h)$ are
 317 conducted for $h \in \{1, \dots, \ell\}$, then $n = \ell \cdot p$ and the vector of observations is:

$$\underline{\mathbf{Y}} = (Y_1(X_{g_1}), \dots, Y_p(X_{g_1}), \dots, Y_1(X_{g_\ell}), \dots, Y_p(X_{g_\ell}))^\top. \quad (1)$$

318 If some observations are incomplete, that is to say some components of
 319 \mathbf{Y}_{g_j} are missing for some j , then $\underline{\mathbf{Y}}$ will be a subvector of $\underline{\mathbf{Y}}$ given in Equation
 320 (1). It means that there may be missing data in the outputs' observations.

321 2.2. Mean and covariances of output variables

322 The originality of the present work is that for a grain g , $\mathbf{Y}(g)$ is defined to
 323 be $\mathbf{Y}(X_g)$, the value of \mathbf{Y} at a random location $X_g \in g$. If the random field
 324 $\{\mathbf{Y}(x) : x \in \chi\}$ and the joint distribution of $\{X_g \in \chi : g \in \mathcal{G}\}$ are known,
 325 then the joint distribution of $\{\mathbf{Y}(g) : g \in \mathcal{G}\}$ can be deduced. And, if one
 326 only knows the moments of order one and cross moments of order two of
 327 $\{Y(x) : x \in \chi\}$ together with the joint distribution of $\{X_g \in \chi : g \in \mathcal{G}\}$,
 328 then one can expect to be able to deduce expectation and cross covariances
 329 of $\{\mathbf{Y}(g) : g \in \mathcal{G}\}$.

330 In the rest of the paper, we assume that the first two moments of
 331 $\{\mathbf{Y}(x) : x \in \chi\}$, $\{X_g \in \chi : g \in \mathcal{G}\}$ and $\{\mathbf{Y}(g) : g \in \mathcal{G}\}$ exist. In the follow-
 332 ing proposition, we show that we can indeed deduce the moments of grains'
 333 outputs.

Proposition 1 (Mean and covariance of $\mathbf{Y}(g)$). *From Definition 2, we derive the following results:*

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

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

(ii) *For any two grains g, g' in \mathcal{G} and any two indices $i, j \in \{1, \dots, p\}$, 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:*

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

In particular,

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

334 *Proof.* (i) is a direct application of the conditional expectation formula
 335 $\mathbb{E}[V] = \mathbb{E}[\mathbb{E}[V|U]]$ where $Y_i(x)$ is the result of conditioning $Y_i(g)$ with X_g .
 336 (ii) is derived from the conditional covariance formula:

$$\text{Cov}[U, V] = \mathbb{E}[\text{Cov}[U, V|W]] + \text{Cov}[\mathbb{E}[U|W], \mathbb{E}[V|W]]$$

337 after conditioning by the joint random vector $(X_g, X_{g'})$ (random variable
 338 X_g). □

339 Note that $\text{Cov}[\mu_i(X_g), \mu_j(X_{g'})] = 0$ when $\mu_i(x)$ is constant over g or g'
 340 or when X_g and $X_{g'}$ are independent. Also note that this framework yields
 341 the expected result that if a grain is restricted to a point, then the output
 342 variables associated with this grain are the same as those associated with the
 343 underlying point.

344 **Example 2.** *For two distinct and finite grains g and g' of cardinalities*
 345 $[g], [g']$, *assuming in this example that X_g and $X_{g'}$ are independent uniform*

346 *random variables, we get:*

$$\begin{aligned}\mu_i(g) &= \frac{1}{[g]} \sum_{x \in g} \mu_i(x) \\ k_{i,j}(g, g') &= \frac{1}{[g][g']} \sum_{(x,x') \in g \times g'} \text{Cov}[Y_i(x), Y_j(x')] \\ k_{i,j}(g, g) &= \frac{1}{[g]} \sum_{x \in g} \text{Cov}[Y_i(x), Y_j(x)]\end{aligned}$$

Remark 1 (Comparison with average – block-to-block covariances). *Previous models using the concept of blocks 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 cumulative distribution function (cdf) of the, possibly discrete, random variable X_g , $i \in \{1, \dots, p\}$. One can check that with this setting, the mean of the mixture $Y_i(g)$ and the average $\bar{Y}_i(g)$ are identical:*

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

Regarding 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,

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

347 *because the independence assumption does not hold any more. As a*
 348 *consequence, $\mathbb{V}[Y_i(g)] \neq \mathbb{V}[\bar{Y}_i(g)]$, even in the specific case where*
 349 *$\forall i, j, g, g', \text{Cov}[\mu_i(X_g), \mu_j(X_{g'})] = 0$. The difference between a mixture and*
 350 *an average is retrieved here: $\mathbb{V}[Y_i(g)] \geq \mathbb{V}[\bar{Y}_i(g)]$.*

351 2.3. Best unbiased linear predictor

352 A Gaussian Process is a collection of random output variables indexed
 353 over points in the input space of explanatory variables, typically denoted
 354 as $Y(\cdot)$. An observation is therefore a random variable $Y(x)$ evaluated at
 355 a given point x , and the covariance between $Y(x)$ and $Y(x')$ is a function
 356 of (x, x') . But we rather consider here an uncertainty on the explanatory
 357 variable, meaning that an observation is modelled as a random field $Y(\cdot)$

358 evaluated at a random location X_g over a given gain g . Thus, one observes
359 a mixture of Gaussian random variables that are not Gaussian any more.
360 Moreover the covariance between $Y(X_g)$ and $Y(X_{g'})$ depends on the joint
361 random variables $(X_g, X_{g'})$. In the previous subsection, some assumptions
362 have been made that are sufficient to be able to compute the covariance be-
363 tween two observations. In the present subsection, it is proved that, given
364 the above defined framework and a learning set of observations, a best linear
365 predictor can be inferred to predict the output features associated with a
366 grain $g \subset \chi$ that has not been observed, given a learning set of observa-
367 tions. Note that the problem amounts to predicting any component of the
368 output variable and that the specific covariance structure resulting from the
369 uncertainty on the explanatory variable requires the development of a new
370 software package, as usual packages such as `DiceKriging` can not fit such a
371 model.

372 Let $\underline{\mathbf{Y}}$ be the vector of observations forming the learning set, and let
373 $g \subset \chi$ be a grain such that for some $i \in \{1, \dots, p\}$, $Y_i(g)$ is to be predicted.

374 Denote:

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

375 where $\mathcal{S}_n^+(\mathbb{R})$ is the set of semi-definite positive, $n \times n$, real matrices.

376 In the following, \mathbf{K} is assumed to be invertible.

377 With a given set of weights $\boldsymbol{\alpha}(g) = (\alpha^1(g), \dots, \alpha^n(g)) \in \mathbb{R}^n$, is associated
378 a linear predictor $M_{\boldsymbol{\alpha}(g)}$:

$$M_{\boldsymbol{\alpha}(g)} = \sum_{j=1}^n \alpha^j(g) Y^j = \boldsymbol{\alpha}(g)^\top \underline{\mathbf{Y}}.$$

379 The optimal weights $\boldsymbol{\alpha}_i(g)$, provided that they exist and are unique, are
380 defined to be those minimizing a quadratic error over all unbiased linear
381 predictors:

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

382 Given the optimal predictor $M_i(g)$, the prediction error and the Kriging

383 (co)variance are denoted as:

$$\begin{aligned}\epsilon_i(g) &:= Y_i(g) - M_i(g) \\ c_{i,j}(g, g') &:= \mathbb{E}[\epsilon_i(g) \epsilon_j(g')] \end{aligned} \quad (2)$$

$$v_i(g) := c_{i,i}(g, g) \quad (3)$$

384 The following proposition gives an optimal predictor that can be com-
385 puted under the minimal assumptions of Proposition 1. Given the first two
386 moments of random variables $\{X_g : g \in \mathcal{G}\}$, all components of $\underline{\boldsymbol{\mu}}$, \mathbf{K} , and
387 $\mathbf{h}_i(x)$ can be computed.

Proposition 2 (Mixture Kriging prediction). *Given a set of observations $\underline{\mathbf{Y}}$, for any $g, g' \in \chi$, and in particular for a single point $g = \{x\}$, for any $i \in \{1, \dots, p\}$, 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 $\underline{\boldsymbol{\mu}} = (0, \dots, 0)^\top$ and $\mu_i(g) = 0$ then*

$$\begin{aligned}\boldsymbol{\alpha}_i(g) &= \mathbf{K}^{-1} \mathbf{h}_i(g) \\ c_{i,j}(g, g') &= k_{i,j}(g, g') - \mathbf{h}_i(g)^\top \mathbf{K}^{-1} \mathbf{h}_j(g') \end{aligned} \quad (4)$$

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

$$\begin{aligned}\boldsymbol{\alpha}_i(g) &= \mathbf{K}^{-1} (\mathbf{h}_i(g) + \lambda_i(g) \underline{\boldsymbol{\mu}}) \\ \text{where } \lambda_i(g) &= \frac{\mu_i(g) - \underline{\boldsymbol{\mu}}^\top \mathbf{K}^{-1} \mathbf{h}_i(g)}{\underline{\boldsymbol{\mu}}^\top \mathbf{K}^{-1} \underline{\boldsymbol{\mu}}} \\ c_{i,j}(g, g') &= k_{i,j}(g, g') - \mathbf{h}_i(g)^\top \mathbf{K}^{-1} \mathbf{h}_j(g') + \lambda_i(g) \lambda_j(g) \underline{\boldsymbol{\mu}}^\top \mathbf{K}^{-1} \underline{\boldsymbol{\mu}} \end{aligned} \quad (5)$$

388 Proof of Proposition 2 is given in Supplementary material Appendix A.

389 Proposition 2 is presented as an algorithm in pseudo-language for Simple
390 Mixture Kriging in Algorithm 1.

391 Assume that $\{\mathbf{Y}(x) : x \in \chi\}$ is a vector-valued Gaussian random field
392 and that each X_g is Dirac distributed for all grains. This last condition holds
393 in particular when each grain is restricted to one singleton point. In this
394 Gaussian case, one retrieves Simple Kriging and Ordinary Kriging predictors,
395 as defined, for example, in Rasmussen and Williams (2006). In this sense,

396 the Mixture Kriging results presented here can be seen as a generalization of
 397 the Kriging interpolation.

398 The above Proposition 2 is valid to predict a single component $Y_i(g)$ of the
 399 output variable $\mathbf{Y}(g)$, but it can be extended to the prediction of $\mathbf{Y}(g)$: the
 400 best linear unbiased predictor of $\mathbf{Y}(g) = (Y_1(g) \dots Y_p(g))^\top$ for the quadratic
 401 error $\mathbb{E} [\|\mathbf{Y}(g) - \mathbf{A}\underline{\mathbf{Y}}\|_2^2]$ is $M_{\mathbf{A}(g)} = \mathbf{A}(g)\underline{\mathbf{Y}}$ where $\mathbf{A}(g)$ is the matrix of
 402 which the i -th row is equal to $\boldsymbol{\alpha}_i(g)^\top$ given by Proposition 2.

403 2.4. Particular cases

404 In this subsection, two important particular cases are explored. The
 405 first one considers the Ordinary Mixture Kriging situation, where the output
 406 variable's expectation is the same everywhere. An estimator of this constant
 407 expectation is derived. The second particular case considers Mixture Kriging
 408 with noisy observations and shows that a nugget effect can be introduced the
 409 same way as for Kriging.

410 **Particular case 1** (Constant mean $\underline{\boldsymbol{\mu}} = \mu_0(1, \dots, 1)^\top$). *Regarding Ordinary*
 411 *Mixture Kriging, assuming that all random variables $Y_i(g)$ have the same*
 412 *unknown expectation μ_0 , and setting $\mathbf{1}_n = (1, \dots, 1)^\top$, Equation (5) simplifies*
 413 *into:*

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

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

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

414 *therefore $\hat{m}(g)$ is an unbiased estimator of μ_0 . \hat{m} can be compared with usual*
 415 *sample mean for independent observations $\underline{\bar{\mathbf{Y}}} = \frac{\mathbf{1}_n^\top \underline{\mathbf{Y}}}{\mathbf{1}_n^\top \mathbf{1}_n}$.*

416 **Particular case 2** (Noisy observations). *Let us consider the case where, for*
 417 *a given $x \in \chi$, we can only observe $\tilde{Y}_i(x) = Y_i(x) + \mathbf{e}_i(x)$ where $\mathbf{e}_i(x)$ is*
 418 *independent from any $Y_j(x')$. We denote the resulting noisy output variables,*
 419 *observations and covariances as follows:*

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

Algorithm 1: Simple Mixture Kriging predictor presented as an algorithm in pseudo-language.

Data:

It is assumed that all grains are discretized and that for any grain g , X_g is uniform. A single output random field $Y(\cdot)$ is observed.

\mathcal{G} : A list of n observed grains $\mathcal{G}_i, i \in \{1, \dots, n\}$, each grain being a table with its points coordinates.

\mathbf{Y} : Observed values $Y^i, i \in \{1, \dots, n\}$, a numeric vector of same length as \mathcal{G} .

$k(\cdot, \cdot)$: Covariance kernel, a function that takes 2 points and returns a positive real number.

g_0 : An unobserved grain to be predicted i.e. a table with its points coordinates.

ϵ : A positive real number giving the nugget effect.

Result:

Optimal weights $\boldsymbol{\alpha}$ for g_0 .

Conditional expectation M of $Y(g_0)$.

Kriging variance $v = \mathbb{V}[Y(g_0) - M]$ (variance of the prediction error).

begin

Fill \mathbf{K} :

for $(i, j) \in \{1, \dots, n\} \times \{1, \dots, n\}, i \geq j$ **do**

if $i = j$ **then**

$\mathbf{K}_{i,i} = \frac{1}{|\mathcal{G}_i|} \sum_{x \in \mathcal{G}_i} k(x, x) + \epsilon$

else $\mathbf{K}_{i,j} = \mathbf{K}_{j,i} = \frac{1}{|\mathcal{G}_i| \times |\mathcal{G}_j|} \sum_{x \in \mathcal{G}_i, x' \in \mathcal{G}_j} k(x, x')$

Fill \mathbf{h} :

for $i \in \{1, \dots, n\}$ **do**

$\mathbf{h}_i = \frac{1}{|\mathcal{G}_i|} \sum_{x \in \mathcal{G}_i, x' \in g_0} k(x, x')$

Get results:

$\boldsymbol{\alpha} = \mathbf{K}^{-1} \mathbf{h}$

$M = \boldsymbol{\alpha}^\top \mathbf{Y}$

$v = \sum_{x \in g_0} k(x, x) - \mathbf{h}^\top \mathbf{K}^{-1} \mathbf{h}$

420 Then the covariance between 2 grains' outputs is:

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

421 Therefore the observations covariance matrix writes:

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

422 And the covariance vector between the observations and a new grain writes:

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

423 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
424 which case \mathbf{K}_e is a diagonal matrix and $\mathbf{h}_{e,i}(g)$ is null as long as g is not
425 among the observed grains.

426 Contrary to Gaussian Process Regression, the prediction cross error
427 $c_{i,j}(g, g')$ defined in Equation (2) is usually not equal to the conditional co-
428 variance $\mathbb{E} [\text{Cov} [Y_i(g), Y_j(g') | \underline{\mathbf{Y}}]]$. However, under certain assumptions, one
429 can prove that if $M_i(g) = \mathbb{E} [Y_i(g) | \underline{\mathbf{Y}}]$, then the cross error can also be viewed
430 as a conditional expectation: $c_{i,j}(g, g') = \mathbb{E} [\text{Cov} [Y_i(g), Y_j(g') | \underline{\mathbf{Y}}]]$. Details are
431 given in Supplementary material Appendix B.

432 3. Illustration

433 3.1. Unidimensional case: rounded inputs

434 A common issue when feeding geo-statistical models with real data is the
435 precision of the input data and its impact on a model's performance. Usual
436 applications of Kriging take this uncertainty into account when increasing
437 output variables' variances by a value that is known as the nugget effect (e.g.
438 Rocas et al., 2021). Precision being a typical case of input data uncertainty,
439 the example below simulates the effect of rounding input values (coordinates

440 in the study space) to the nearest units. Let us consider a one-dimensional,
 441 centred Gaussian random field $Y(x)$, $x \in [1, 10]$ of constant variance. Let
 442 us assume that this field is observed at some points for which coordinates
 443 are rounded to the nearest unit, i.e., for 2 input values $x_1, x_2 \in]0.5, 1.5]$,
 444 the observer sees the same value $\tilde{x}_1 = \tilde{x}_2 = 1$. For a Kriging model, these
 445 are multiple observations of the same point, and it is necessary to introduce
 446 a nugget effect in the model for the observations' covariance matrix to be
 447 invertible. This nugget effect simulates an uncertainty on the output values,
 448 while the uncertainty is really on the input values. It rather makes sense to
 449 describe those input values as random positions $\tilde{x}_{1,g}$ and $\tilde{x}_{2,g}$ in $g =]0.5, 1.5]$
 450 instead of deterministic $\tilde{x}_1 = \tilde{x}_2 = 1$. Then, we can model the observed
 451 objects as mixture distributions and fit a Mixture Kriging model. Let us
 452 compare both approaches.

453 Using the `geoR` package in the R language, we simulate a 1-dimensional
 454 random field realization with a Gaussian covariance kernel. The specific pa-
 455 rameters are detailed in Table 3. x is discretized between 0 and 10 with step
 456 0.05. We pick 8 points for observations as listed in Table 4. These observa-
 457 tions are plotted on Figure 3. Observations $\{o1, o2, o6\}$ form the learning set,
 458 observations $\{o4, o5, o7\}$ form the validation set, and observations $\{o3, o8\}$
 459 form the test set. These sets of observations are deliberately very small so
 460 as to represent the Mixture Kriging's behavior in a readable graphic.

Underlying field		Model properties				Validation	Total
Variance	Range	Model	Variance	Nugget	Range	MSE	MSE
1	4	Kriging	1	10^{-9}	4	0.037	1.14
1	4	Mixture Kriging	1	0	4	0.027	1.18

Table 3: Parameters and performances of fitted models in the case of observations with rounded inputs. Note that the nugget effect for Kriging is the result of an optimization process. For Mixture Kriging, nugget is null by design. Validation MSE: Mean Squared Error on validation set. Total MSE: Mean Squared Error on the complete interval $[0, 10]$.

461 The Kriging model (Figure 3 left) has repeated observations for $x = 1$
 462 and $x = 3$. The learning set is used to fit a family of models with the same
 463 kernel parameters as those used for simulation plus a nugget effect among
 464 $(10^{-i})_{i \in \{1, \dots, 10\}}$. The nugget effect yielding the smallest mean squared error
 465 (MSE) on the test set is selected. A new model is fitted with both learning
 466 and test sets using the same kernel and the previously selected nugget ef-
 467 fect. This model is applied to compute a validation MSE and a total MSE

Set	Label	Input			Output
		Underlying x (true value)	Rounded x (for Kriging)	Grain (for Mixture Kriging)	y
Learning	$o1$	0.55	1	$g_1 =]0.5, 1.5]$	0.923
Learning	$o2$	0.85	1	$g_2 =]0.5, 1.5]$	1.005
Validation	$o3$	1.65	2	$g_3 =]1.5, 2.5]$	1.127
Test	$o4$	3.00	3	$g_4 =]2.5, 3.5]$	0.946
Test	$o5$	3.45	3	$g_5 =]2.5, 3.5]$	0.801
Learning	$o6$	7.20	7	$g_6 =]6.5, 7.5]$	0.337
Test	$o7$	9.40	9	$g_7 =]8.5, 9.5]$	0.884
Validation	$o8$	9.70	10	$g_8 =]9.5, 10]$	0.908

Table 4: Observations of the simulated Gaussian random field.

468 computed on all points in $[0, 10]$. The variance of the prediction error is also
469 predicted using the formula given in Proposition 2.

470 Regarding Mixture Kriging (Figure 3 right), grains $g_1 = [0.5, 1.5[$ and
471 $g_3 = [2.5, 3.5[$ are observed twice each while the other grains are observed
472 once each. The Mixture Kriging model can handle repeated observations
473 by design. Uncertainty on the input is resulting from the random position
474 that generates the observation. The grain covariances are computed from
475 the point covariances as detailed in Proposition 1. The random positions
476 $(X_{g_i})_{i \in \{1, \dots, 8\}}$ are assumed to be uniform on the points of the associated grains.
477 Both the learning set and the test set are used to fit a model with the same
478 kernel parameters as for simulation and with no nugget effect. Validation
479 MSE and total MSE are computed for comparison with Kriging.

480 In this case, the mean prediction is almost the same for both models. But
481 Kriging variance (visible on the ribbons in Figure 3) differs. By construction,
482 Simple Kriging is supposed to interpolate observations exactly, resulting in a
483 very small Kriging variance near observations. Too many observations may
484 be outside the confidence band. If one increases the nugget effect on the
485 Simple Kriging model, mean predictions move towards 0 and their range is
486 reduced. Therefore, with a large nugget effect, one ends up with a nearly
487 constant mean prediction and a large Kriging variance. Mixture Kriging
488 takes into account the input uncertainty and predicts a significant Kriging
489 variance even near observations improving the coverage without any nugget
490 effect.

491 In this very simple example, the reader may be surprised that both
492 Kriging and Mixture Kriging yield remarkably good predictions. However,

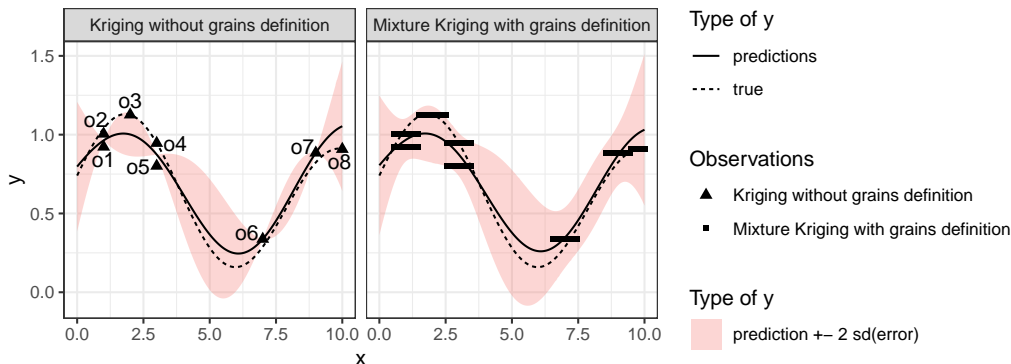


Figure 3: *Rounded inputs*. **Left and right:** The dashed line shows the underlying simulated random field. The solid line labelled “predictions” shows the fitted model mean prediction (see Table 3). The ribbon shows an interval of radius twice the root square of the estimated prediction error variance. **Left:** *Kriging model*. Triangular points show observations. **Right:** *Mixture Kriging*. Horizontal line segments show observations. See Table 4 for more details about observations.

493 the prediction error value represented by the ribbon’s height is important
 494 as compared with the predicted values. This means that if the underlying
 495 output is also noisy, error can quickly increase. This is the reason why, in
 496 real life, one needs much more observations to learn from, see Subsection 3.3.

497 3.2. Unidimensional case: grains of varying size

498 Imagine a company that wants to measure some performance indicator
 499 for manufactured objects that are produced according to certain design spec-
 500 ifications. The design is denoted as x ; it belongs to a set χ of permissible
 501 values, and $\mathbf{Y}(x)$ is the performance indicator. For instance, \mathbf{Y} can measure
 502 the lift of an aircraft wing depending on some shape parameter x . Because of
 503 some unavoidable manufacturing precision issues, the manufactured object’s
 504 characteristics do not match the design’s specifications exactly. This uncer-
 505 tainty about the manufactured object induces some uncertainty about the
 506 performance indicator. Thus, the constructed design can be viewed as a ran-
 507 dom vector X_{g_x} , taking values in some tolerance set $g_x \subset \chi$ around the design
 508 $x \in \chi$. When testing some designs x_1, \dots, x_n , the industry observes perfor-
 509 mances $\mathbf{Y}(g_1), \dots, \mathbf{Y}(g_n)$. Measuring both the expectation and the variance
 510 of $\mathbf{Y}(x)$ for each permissible design $x \in \chi$ is one method to find the best de-
 511 sign, but this can be costly, so fitting an interpolation model with the set of k
 512 observations is preferable. In this setting, for the sake of simplicity, we assume
 513 that $\mathbf{Y}(x)$ is conditioned by observations $\{\mathbf{y}(x_i) = \sin(x_i^2) : i \in \{1, \dots, n\}\}$.

514 In this case, we assume that the precision associated with a design x_i is an
515 interval centred on x . The real characteristic of the object having perfor-
516 mance $\mathbf{y}(x_i)$ is a random value in this grain, which is assumed to be uniform
517 on all points of the grain.

518 We compare 3 models:

- 519 • P_1 : The manufactured object is produced exactly according to the
520 design, the precision interval is restricted to a point.
- 521 • P_2 : The precision is the same for all designs, the associated interval is
522 of fixed measure.
- 523 • P_3 : The larger is x , the larger is the uncertainty about the manufac-
524 tured object, which means that intervals' measures are growing with
525 the design x .

526 All three models have a null nugget effect and a Gaussian kernel having
527 for variance parameter the overall variance of \mathbf{y} on $\chi = [0, 4]$. The range
528 parameter is optimized by minimizing the mean squared error between \mathbf{y}
529 and point predictions on χ (see Table 4). When grains are restricted to
530 points (Figure 5 top), we get the usual results on Simple Kriging, in partic-
531 ular, predicted values are exactly interpolating observations. When grains
532 are intervals of the same size (Figure 5 middle), predicted values are not
533 interpolating any more; predicted error is not null on the grains but far from
534 the grains, it is smaller than in the previous case. In the bottom figure, the
535 greater is x , the greater the uncertainty about the manufactured object as
536 compared to design. The predicted error (ribbon) is increasing with the grain
537 diameter. Overall, it is important to note that the Mixture Kriging model
538 accounts for the randomness of input values without any nugget effect. This
539 eliminates the adverse consequences of a nugget effect that could otherwise
540 shrink mean predictions towards zero.

541 In the previous example, a very small set of observations was enough to
542 make very good predictions. In the present one, the situation is different
543 because observations are not drawn from a Gaussian random field but from
544 a deterministic function. This underlying function is modelled as a noisy
545 random field. Therefore, the Kriging error is greater than in the previous
546 case. A potential extension of this illustration would be to optimize both the
547 range and the nugget effect, but the purpose here is to visualize the effect of
548 the uncertainty on the input and not the output.

Set of grains	Model properties			
	Variance	Nugget	Range	Exact interpolation
P_1 : Grains are singletons	0.36	0	0.3	Yes
P_2 : Grains are of equal measure	0.36	0	0.4	No
P_3 : Grains are of increasing measure	0.36	0	0.3	No

Figure 4: Properties of models P_1, P_2, P_3 . Range is an optimal value so as to minimize mean squared error.

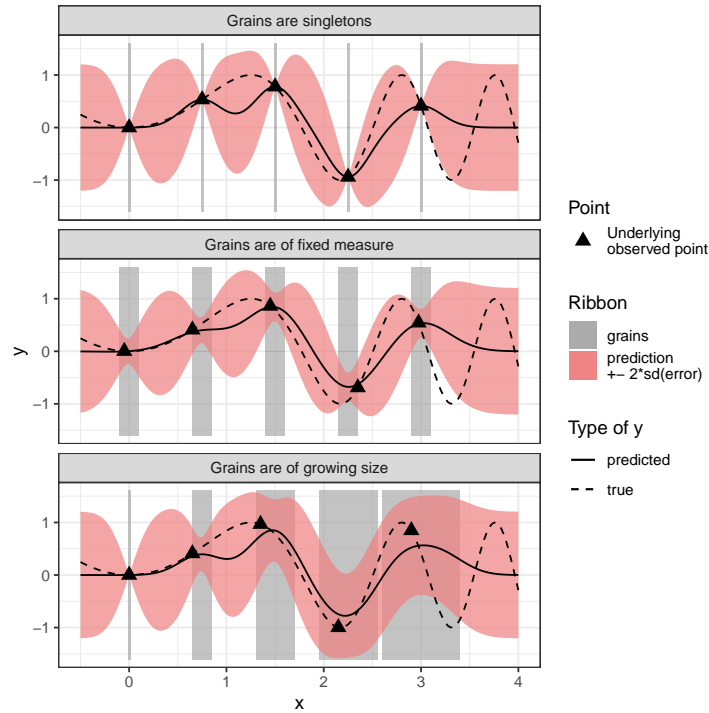


Figure 5: *Mixture Kriging and grain sizes*. The dashed line represents $\mathbf{y}(x)$. The solid line is the mean prediction. The ribbon shows an interval centred on the mean prediction, of radius twice the square root of the predicted error variance. Grey vertical columns show the grains as x intervals. Black triangles show the underlying observed point (observed X_g and associated output value).

549 3.3. Energy Performance Certificate (EPC) prediction

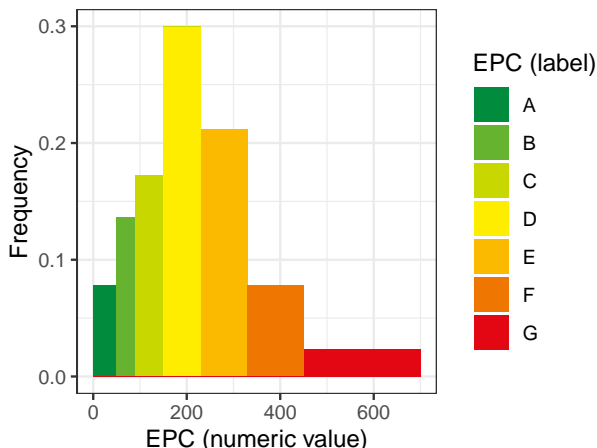


Figure 6: Bar plot of EPC labels frequencies among all EPCs collected in France between 2014 and 2021. Classes A, B, F, G are rare while classes C, D and E are frequent.

550 Let us now address the EPC prediction problem, keeping in mind that
 551 an Energy Performance Certificate (EPC) is given as an energy consumption
 552 in $kWh/m^2/year$. The observed distribution of this energy consumption is
 553 provided in Figure 6. One considers a model for which χ is a city viewed as a
 554 2-dimensional space with latitude and longitude as coordinates after proper
 555 projection, \mathcal{G} is the set of plots, and a point in χ is associated with a given
 556 floor square meter of a building on the plot. A floor square meter is regarded
 557 here as a granule and not as a set of points in χ . This would not make sense
 558 since, for a multi-storey building, there are more floor square meters than
 559 the building's footprint area. $x \in \chi$ is therefore a reference point for this
 560 floor square meter the same way a point would be used to locate a citizen in
 561 a city (see Example 1). $Y(x)$ is the areal energy consumption in x , typically
 562 the EPC of the dwelling to which belongs the floor square meter represented
 563 by x . Then an EPC in the database is the observed energy efficiency rating
 564 associated with one unknown point among those located on the plot indicated
 565 by the address. Therefore, for a certain plot g , this EPC is an observation
 566 of $Y(X_g)$.

567 EPC is given as a numeric energy consumption per square meter and
 568 per year. This energy consumption is associated with a letter ranging
 569 from A to G. A and B label the most energy-saving dwellings (less than
 570 $90kWh/m^2/year$). F and G label the most consuming dwellings (more than

571 330kWh/m²/year). We want to model a situation where we observe EPC
 572 with uncertainty on the location of the observed dwelling on the land plot,
 573 where it lies, and where the observed dwelling can not be distinguished among
 574 all the dwellings of this land plot. We also want to predict an EPC numeric
 575 value at the whole land plot level, that is, for the set of dwellings it contains.

576 As can be seen in Figure 6, observations are strongly unbalanced, mean-
 577 ing that labels A, B, F, and G are rarely observed while labels C, D, and E
 578 are very common. As a result, labels A, B, F, and G are difficult to predict,
 579 although they are more interesting for decision-makers. Therefore, we intro-
 580 duce the Balanced Accuracy (BA) criterion. It is an asymmetric performance
 581 measure that focuses on good results (Gösgens et al., 2021) and it gives the
 582 same weight to each class. Denoting n_ℓ the number of observations with label
 583 ℓ and $n_{\hat{\ell},\ell}$ the number of predictions $\hat{\ell}$ with true label ℓ (true predictions of
 584 label ℓ), the balanced accuracy is given by the formula:

$$BA = \frac{1}{7} \sum_{\ell \in \{A, \dots, G\}} \frac{n_{\hat{\ell},\ell}}{n_\ell}$$

585 Given a real random variable X and F_X its cdf, supposed to be invertible.
 586 Let $H(X) := F_N^{-1} \circ F_X(X)$ where F_N is the standard Gaussian distribution
 587 cdf. H is invertible, and $H(X)$ follows a standard Gaussian distribution by
 588 the probability integral transform theorem. Using H we normalize input and
 589 output variables.

590 Let us consider the model M_1 such that:

- 591 • χ is the territory of an urban area in the French city of Angers in a
 592 3-dimensional space where coordinates represent the image through H
 593 of the construction year, the latitude, and the longitude.
- 594 • A random field $Y(x)$ is defined on χ . It represents the image through
 595 H of the energy consumption per square meter and per year at x .
- 596 • A grain g is defined as a set of points in a 3-dimensional space χ . A
 597 grain represents a land plot. Each point represents a square meter of
 598 living area. It has 3 coordinates. \mathcal{G} denotes the set of all grains.
- 599 • For any grain $g \in \mathcal{G}$, the random variable X_g is the uniform law on
 600 the points of g . It represents the uncertainty on the observations' lo-
 601 cation. On g , the output variable is defined as: $Y(g) = Y(X_g)$. By
 602 construction, Y is centred.

- A vector of observations of n distinct grains is given and denoted as \underline{Y} .

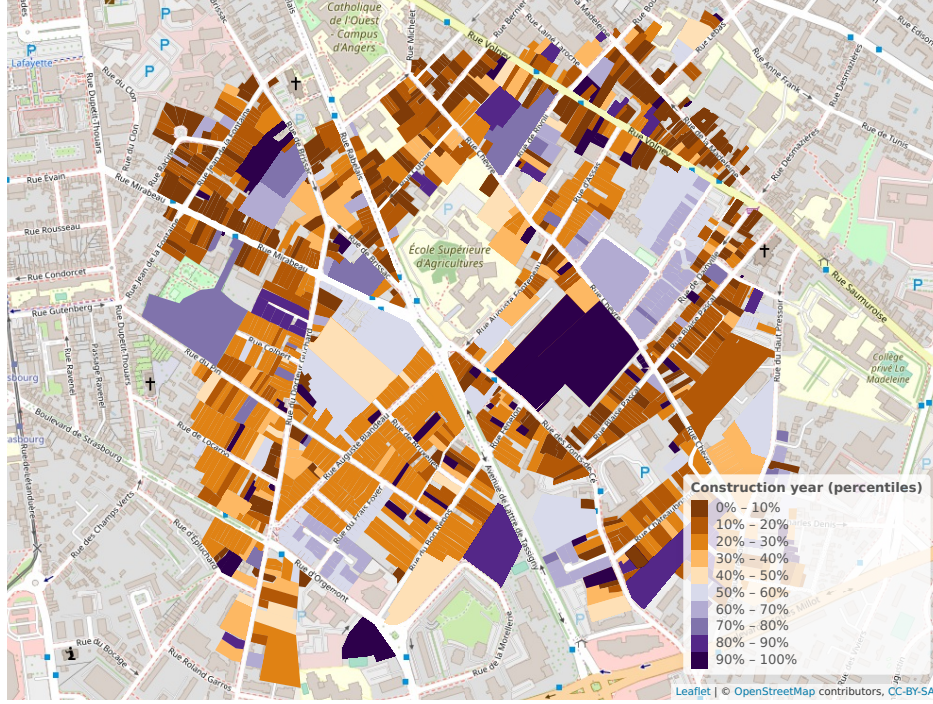


Figure 7: An urban area in Angers: latitude is the vertical dimension, longitude is the horizontal dimension, and construction year is given by the colour. The side of the square is 1km. Construction years range from 1340 (first percentile) to 2019 (last percentile).

604 \mathcal{G} is mapped in Figure 7. Note that the grains seem to be disjoint, but
 605 they are not due to overlaps in the construction year dimension. The set of
 606 observations is represented in Figure 8.

607

608

For this model, the following assumptions are made:

609

610

611

612

- For any two distinct grains g, g' , random variable X_g is independent from $X_{g'}$.
- For any two points x, x' , the covariance between $Y(x)$ and $Y(x')$ is following a Matérn $3/2$ model:

$$\text{Cov}[Y(x), Y(x')] = \sigma^2 \left(1 + \sum_{i=1}^3 \frac{|x_i - x'_i|}{\theta_i} \right) \exp \left(- \sum_{i=1}^3 \frac{|x_i - x'_i|}{\theta_i} \right)$$

where $U = (\sigma^2, \theta_1, \theta_2, \theta_3) \in]0, 1] \times]0, +\infty[^3$

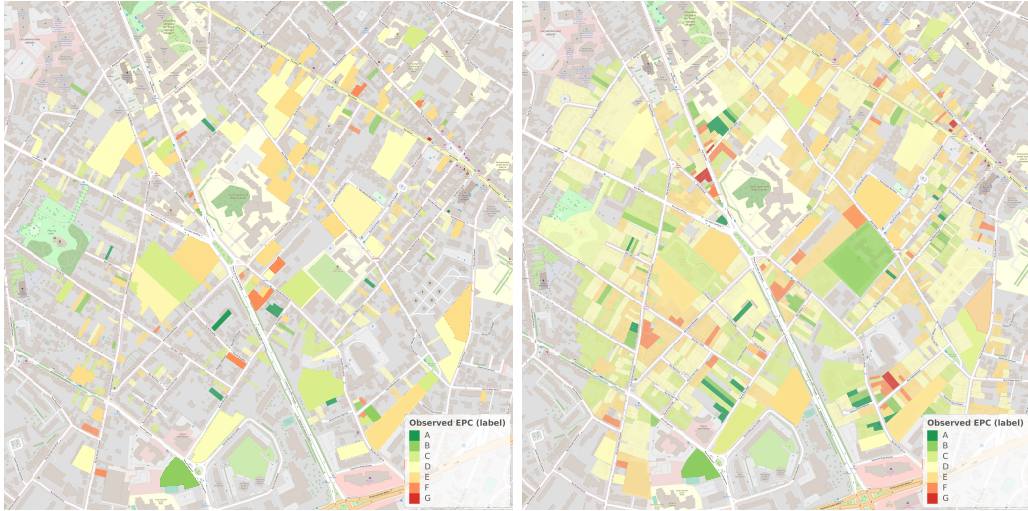


Figure 8: Left: Map of the 365 observations. Right: Map of all predicted values (labels derived from Mixture Kriging means). Each colour represents a label associated with a numeric value. See also Figure 1.

613 σ^2 is called the variance coefficient, and $\Theta = (\theta_1, \theta_2, \theta_3)$ are the
 614 length scale coefficients. Note that no nugget effect is required because the
 615 model takes into account the spatial uncertainty of the input by construction.
 616

617 The Mixture Kriging predictor described in subsection 2.3 is used to
 618 predict energy consumption at the plot level. It can be proved that, without
 619 the nugget effect, the mean prediction, in the case of a 1-dimensional output,
 620 does not depend on σ^2 (the proof is simply deduced from the fact that for
 621 an invertible matrix A , we have $(\lambda A)^{-1} = \lambda^{-1} A^{-1}$). σ^2 is therefore set to
 622 1. Θ is chosen so as to maximize the BA criterion of the predicted labels
 623 derived from the predicted energy consumptions. BA is computed using
 624 leave-one-out cross-validation. Note that the leave-one-out cross-validation
 625 predictor that is derived from Proposition 2 is also linear and optimal for
 626 quadratic error. A code has been developed in the R language to implement
 627 Mixture Kriging.

628
 629 So as to assess the effect of balanced accuracy on the optimum,
 630 we also consider a model $M1'$, which is the same as $M1$ except that
 631 parameters are assessed optimizing the accuracy. The accuracy is the total
 632 number of labels correctly predicted divided by the number of predictions.

633

634 Let us now consider the Kriging model $M2$ to compare performances
635 with the Mixture Kriging model $M1$. $M2$ has the same properties as $M1$
636 presented above, except that:

- 637 • Grains are singletons. A grain $g = \{x^1, \dots, x^q\}$ is replaced by a point x of
638 coordinates the minimum construction year and the mean latitude and
639 longitude values. Note that it is assumed that the year of construction
640 of the eldest building portion is the most meaningful information for
641 prediction. This makes sense, especially because the oldest part of a
642 building is usually also the largest one.
- 643 • A nugget effect σ_e^2 has to be introduced so as to have a smooth predic-
644 tor:

$$\mathbb{V}[Y(x)] = \sigma^2 + \sigma_e^2 .$$

645 For $M2$, the Kriging predictor is used. $V = (\sigma^2, \theta_1, \theta_2, \theta_3, \sigma_e^2)$ is chosen
646 so as to maximize BA, the same way as for $M1$. The standard R package
647 `DiceKriging` is used for prediction.

648

649 There are 365 observations on the given territory. The best parameters
650 are estimated by optimizing the performance indicator, Balanced Accuracy or
651 Accuracy, computed by leave-one-out cross validation. All models $M1$, $M1'$
652 and $M2$ are optimized with the genetic algorithm provided by R package
653 `ga` parametrized with population size 50, elitism 5, maximum number of
654 iterations 100, maximum number of iterations without improvement 100.
655 Other parameters are left as default.

656 With regards to the optimal parameters in Table 5, length scale paramete-
657 rs are smaller in $M1$ than in $M2$, meaning that $M1$ prediction is influenced
658 by fewer neighbours than $M2$. The nugget effect found for $M2$ is small. As
659 for the optimal performances in Table 6, $M1$ reaches a larger BA than $M2$
660 by 37%. However, $M1$ has lower performances on other indicators with a
661 difference of approximately 10%. The range of all 365 mean predictions with
662 $M1$ is 150% larger than with $M2$. These figures are better understood by
663 examining the confusion matrices in Tables 7 and 8. Indeed, the percentage
664 of large errors (represented by the red area) is 3% with model $M1$ and 0.5%
665 with model $M2$. We know that large errors have an important impact on

Model	ϵ^2	σ^2	θ_1	θ_2	θ_3
Mixture Kriging ($M1$)	0.00^*	1.00^*	0.28	0.44	1.22
Mixture Kriging ($M1'$)	0.00^*	1.00^*	0.93	0.78	0.91
Kriging ($M2$)	0.02	0.53	0.98	0.82	1.49

*: These parameters are treated as constant parameters.

Table 5: Optimal parameters for $M1$ and $M2$.

Model	EPC int.		EPC num.				
	BA	Acc.	MAE	RMSE	MAE	RMSE	Range
Mixture Kriging $M1$	0.26	0.40	0.93	1.37	78.93	106.16	6.66
Mixture Kriging $M1'$	0.21	0.42	0.93	1.38	79.46	108.55	6.54
Kriging $M2$	0.19	0.38	0.85	1.22	72.22	92.98	2.59

EPC int.: Energy Performance Certificate treated as an integer: 1 for A, ..., 7 for G.

EPC num.: Energy consumption expressed in $kWh/m^2/year$.

BA: Balanced Accuracy. Acc.: Accuracy.

MAE: Mean Absolute Error. Range: Variance of the predicted values ($\times 10^3$)

RMSE: Root Mean Squared Error. viewed as a measure of the predictions' range.

Table 6: Optimal performances achieved by $M1$, $M1'$ and $M2$. For each indicator, best value is indicated in bold font.

666 MAE and RMSE. However, the percentage of true labels A and B that are
667 predicted as A or B is 25% with $M1$ and 10% with $M2$. For labels F and G,
668 these figures are 16% and 0% respectively. This information is valuable for
669 decision-makers seeking to identify energy-efficient and/or energy-intensive
670 dwellings.

671 These results suggest that Mixture Kriging ($M1$, $M1'$) predictions have
672 an improved range as compared to Kriging ($M2$): the range of mean pre-
673 dictions by Mixture Kriging is greater than by Kriging. This allows better
674 predictions for extreme labels A, B, F, and G. Despite having fewer param-
675 eters (ϵ^2 and σ^2 are regarded as constants), Mixture Kriging improves the
676 BA, although it also leads to more frequent large errors. Mixture Kriging ac-
677 counts for uncertainty in the input data, eliminating the need to add a nugget
678 effect. In this example, it avoids grouping predictions near the mean value
679 (shrinkage) and yields a better BA as compared to Kriging that requires the
680 introduction of a nugget effect.

681 Among Mixture Kriging models, as expected, $M1$ has a better Balanced
682 Accuracy than $M1'$, and $M1'$ has a better Accuracy than $M1$. Other indi-

True values	Predicted values							True values	Predicted values						
	A	B	C	D	E	F	G		A	B	C	D	E	F	G
A	2	1	3	2	2	0	0	A	1	0	3	5	1	0	0
B	1	3	1	9	2	2	0	B	0	2	1	11	4	0	0
C	1	3	25	26	15	4	0	C	0	1	13	48	12	0	0
D	3	5	21	80	33	5	1	D	2	1	19	94	32	0	0
E	4	2	12	36	36	5	1	E	0	1	9	56	30	0	0
F	0	3	2	4	5	3	0	F	1	0	2	11	3	0	0
G	0	0	0	1	1	0	0	G	0	0	1	1	0	0	0

Table 7: Confusion matrix of $M1$ predictions.

Table 8: Confusion matrix of $M2$ predictions.

True values	Predicted values						
	A	B	C	D	E	F	G
A	0	2	2	5	1	0	0
B	1	0	3	9	3	2	0
C	2	2	23	29	14	4	0
D	1	6	17	91	28	2	3
E	1	6	14	31	36	6	2
F	1	0	3	8	2	3	0
G	0	0	1	0	1	0	0

Table 9: Confusion matrix of $M1'$ predictions

	True	Predicted
A	10	11
B	18	17
C	74	64
D	148	158
E	96	94
F	17	19
G	2	2

Table 10: Distribution of labels in $M1$

683 cators are very similar, let alone the smaller variance of $M1'$'s predictions.
684 Optimizing parameters based on Balanced Accuracy forces the model to pre-
685 dict more often labels A, B, F, and G so that the distribution of predicted
686 labels is very close to the distribution of observed labels as can be seen in
687 Table 10. In our case, the confusion matrices show that this effect is positive
688 for labels A and B, as more true A or B are predicted as A or B. But the
689 effect of balanced accuracy does not bring benefits for labels F and G, on the
690 contrary, it has a tendency to predict more F and G where the true label is
691 D or E. A possible explanation for this moderate benefit of introducing the
692 balanced accuracy is that we are missing some information. The moderate
693 size of observations (365 individuals) makes it difficult for a model to discrim-
694 inate between rare labels and frequent labels. For instance, there are only 2
695 observed G labels. One can expect a model learning from a larger number of
696 observations to perform better. Moreover, in an area where buildings are

697 old for instance, our model cannot distinguish a building that has never been
698 renovated from the others. It may be useful in further studies to introduce
699 more variables, such as a comfort level. However, as discussed below, the
700 proposed model is quite heavy in terms of computation resources; therefore,
701 scaling up or adding variable has an important computational cost.

702 4. Discussion and conclusion

703 Since the discovery of Kriging, the issue of learning from and predict-
704 ing areal data has been a concern. Proposed models have mainly assumed
705 that the output variable at the areal level is the mean of the point outputs,
706 which has proven helpful in various fields such as mining, climatology, or
707 satellite imaging, where averaging makes sense for interpretation and where
708 blocks tend to have similar shapes and sizes. However, in other fields such
709 as agriculture or social studies, blocks can have varying shapes or sizes, and
710 averaging is not always the most meaningful interpretation. In these cases,
711 problems like the Modifiable Areal Unit Problem (MAUP), the ecological
712 inference problem, and the variance reduction due to averaging can become
713 challenging to solve. Over the past few decades, researchers have been de-
714 veloping methods to assess and/or correct the MAUP effect (Briz-Redon,
715 2022). Modifying territory partitioning (Li et al., 2009) is also an effective
716 solution for addressing variance reduction problems, but it is not always pos-
717 sible. Both Kriging and block-Kriging incorporate uncertainties on input
718 and/or output values through the addition of a nugget effect to variances,
719 thereby simulating the addition of a white noise to the output variables. This
720 transformation smooths predicted values but also shrinks them; the range be-
721 tween minimal and maximal predicted values is reduced, thus degrading the
722 prediction quality of values that are particularly large or particularly small.

723 The availability of new datasets with uncertainty on the inputs (uncertain
724 positions) and where averaging is not a meaningful interpretation has driven
725 us to seek a novel method of spatial interpolation. We have introduced a
726 new element in the model that is a random input value. It has been found
727 that resulting mixture distributions can be interpolated optimally, and the
728 resulting Best Linear Unbiased Predictor (BLUP) requires only the first 2
729 moments of the prior random field and a spatial covariance function. This
730 model can learn from and predict outputs associated with grains of any shape,
731 size, or cardinality. Even single points are acceptable. The term “grain” has
732 been introduced to describe these objects.

733 The new model called Mixture Kriging is still consistent with Kriging in
734 the sense that Kriging is a special case of Mixture Kriging where grains are
735 restricted to singletons. However, Mixture Kriging generates a mean pre-
736 diction range that is not impacted by the grain’s shape or size under usual
737 conditions. As a consequence, there is no reduction in the mean prediction’s
738 range due to this factor. If the output variable’s variance is the same every-
739 where at point level, then it is also the same as the output variable’s variance
740 at grain level, meaning that there is no variance reduction either. Similarly,
741 if the covariance between the output variable of interest and another output
742 variable is the same everywhere at the point level, then it will also be the
743 same as the covariance at the grain level, regardless of the grain’s shape.
744 This implies that this model has no measurable MAUP effect in the sense of
745 Briz-Redon (2022).

746 Without any MAUP effect, the Mixture Kriging approach is able to han-
747 dle multi-scale data. We hope that this can help handling datasets coming
748 from multiple sources in the same model. This model can potentially be
749 used to fit ecological data or social data. For instance, on a global scale,
750 the Intergovernmental Panel on Climate Change (IPCC) studies planetary
751 boundaries on water based on gridded data, but for the study of specific
752 territories, studies are commonly done on watersheds. The Mixture Kriging
753 model has the ability to combine these two scales of study in the same model
754 to benefit from both global and local studies.

755 The main computational distinction between block-to-block Kriging and
756 Mixture Kriging lies in the method of computing the observations variance
757 and the covariance between covariates associated with the same grain. This
758 results mainly in the diagonal of the observations covariance matrix being
759 greater than what is found with Kriging. This is precisely the sought ef-
760 fect when introducing supplementary noise on the outputs (nugget effect) in
761 Kriging for smoothing predictions. This explains why Mixture Kriging has
762 smooth predictions but with limited shrinkage, hence a good performance
763 with Balanced Accuracy. In practical applications, Mixture Kriging is there-
764 fore designed to handle data with uncertainty on the input values without
765 introducing the nugget effect.

766 Regarding computational differences, it should also be noted that Mixture
767 Kriging (like block-to-block Kriging) has a higher computational cost than
768 Kriging, this cost is growing like the squared value of the density of points
769 in the grains. This is an important limitation of the model. For instance,
770 in the models $M1$, $M1'$ and $M2$ presented in Subsection 3.3, there are 395

771 observations. The Kriging model $M2$ requires $365 \times 366/2 = 66,795$ covari-
772 ances to be computed. But the Mixture Kriging models $M1$ and $M1'$ require
773 to compute 3,770,500,618 point-to-point covariance in order to compute the
774 66,795 covariances between grains. Scaling up the model may, therefore, be
775 difficult. This computational complexity is highly dependent on the defini-
776 tion of the random position X_g for each grain and on its discretization. For
777 the above models, X_g is supposed to be uniform for all grains, and the num-
778 ber of discretized points is the number of square meters of living space on
779 the grain. But any new model based on Mixture Kriging requires an appropri-
780 ate definition of these random variables, depending both on the grains'
781 geometries and on the studied output variable(s). Another limitation of the
782 model is the difficulty of assessing its parameters, especially the range. It is
783 difficult to compute a variogram because there is no natural definition of a
784 distance between grains. Estimating the range is also possible by minimizing
785 an error measure, but this process requires computing numerous different
786 models, which is costly, as mentioned above.

787 Keeping in mind its limitations, this new approach opens the way for
788 implementing Mixture Kriging models with new datasets that have been im-
789 possible to fit in the usual Kriging framework or with usual Kriging software
790 packages. In particular, datasets that inform about granules that are un-
791 certainly defined, such as dwellings, buildings, streets, human persons, and
792 households. It can also be used for datasets informing about granules, which
793 should have deterministic shapes or positions in the input space, but come
794 with numerical uncertainty such as measure precision, rounding effect, obser-
795 vations' aggregations, or observations' anonymization. Moreover, the model
796 can handle multivariate outputs, even if some output components are miss-
797 ing in the observations. Encouraging results have been found when studying
798 the prediction of Energy Performance Certificates (EPC). Results show that
799 Mixture Kriging can be useful to improve the prediction of values far from the
800 average and, in our case, to improve the detection of energy-saving homes.
801 Future studies should test the upscaling feasibility of the already developed
802 model and the benefits of using covariates. We also study the possibility of
803 developing a similar model with Universal Kriging.

804 **Acknowledgements**

805 The authors acknowledge support from the URBS enterprise, [www.urbs.](http://www.urbs.fr)
806 [fr](http://www.urbs.fr). They thank in particular Maximilien Brossard for careful reading and
807 constructive comments.

808 This research was jointly supported by Mines Saint-Etienne graduate
809 engineering school and research institute ([https://www.mines-stetienne.](https://www.mines-stetienne.fr/en/)
810 [fr/en/](https://www.mines-stetienne.fr/en/)), URBS enterprise (<https://www.imope.fr/>) and French National
811 Agency for Research and Technology (<https://www.anrt.asso.fr/fr>).

812 **Bibliography**

- 813 Aldworth, W. J. K. (1998). *Spatial prediction, spatial sampling, and mea-*
814 *surement error*. Doctor of Philosophy, Iowa State University.
- 815 Ali, U., Shamsi, M. H., Bohacek, M., Hoare, C., Purcell, K., Mangina, E.,
816 and O'Donnell, J. (2020). A data-driven approach to optimize urban scale
817 energy retrofit decisions for residential buildings. *Applied Energy*, 267.
- 818 Baker, E., Challenor, P., and Eames, M. (2021). Future Proofing a Building
819 Design using History Matching Inspired Level-set Techniques. *Journal of*
820 *the Royal Statistical Society Series C: Applied Statistics*, 70(2):335–350.
- 821 Ballarini, I., Corrado, V., Madonna, F., Paduos, S., and Ravasio, F. (2017).
822 Energy refurbishment of the Italian residential building stock: energy and
823 cost analysis through the application of the building typology. *Energy*
824 *Policy*, 105:148–160.
- 825 Briz-Redon, A. (2022). A Bayesian shared-effects modeling framework to
826 quantify the modifiable areal unit problem. *Spatial Statistics*, 51.
- 827 Comber, A. and Zeng, W. (2019). Spatial interpolation using areal fea-
828 tures: A review of methods and opportunities using new forms of
829 data with coded illustrations. *Geography Compass*, 13(10). _eprint:
830 <https://onlinelibrary.wiley.com/doi/pdf/10.1111/gec3.12465>.
- 831 Cressie, N. A. (1993). *Statistics for spatial data revised edition*. Wiley se-
832 ries in probability and mathematical statistics. Applied probability and
833 statistics., New York, john wiley & sons inc. edition.
- 834 Godoy, L. d. C., Prates, M. O., and Yan, J. (2022). An unified framework for
835 point-level, areal, and mixed spatial data: the Hausdorff-Gaussian Process.
- 836 Goovaerts, P. (2008). Kriging and Semivariogram Deconvolution in the Pres-
837 ence of Irregular Geographical Units. *Mathematical Geology*, 40(1):101–
838 128.
- 839 Gösgens, M., Zhiyanov, A., Tikhonov, A., and Prokhorenkova, L. (2021).
840 Good Classification Measures and How to Find Them. In Ranzato, M.,
841 Beygelzimer, A., Dauphin, Y., Liang, P. S., and Vaughan, J. W., editors,
842 *Advances in Neural Information Processing Systems*, volume 34, pages
843 17136–17147, New York. Curran Associates, Inc.

- 844 Gotway, C. and Young, L. (2002). Combining Incompatible Spatial Data.
845 *Journal of the American Statistical Association*, 97:632–648.
- 846 Isaaks, E. H. and Srivastava, R. M. (1989). *An Introduction to Applied*
847 *Geostatistics*. Oxford University Press, New York.
- 848 Jin, Y., Ge, Y., Wang, J., Heuvelink, G., and Wang, L. (2018). Geographi-
849 cally Weighted Area-to-Point Regression Kriging for Spatial Downscaling
850 in Remote Sensing. *Remote Sensing*, 10.
- 851 Kerry, R., Goovaerts, P., Smit, I. P., and Ingram, B. R. (2013). A comparison
852 of multiple indicator kriging and area-to-point Poisson kriging for mapping
853 patterns of herbivore species abundance in Kruger National Park, South
854 Africa. *International journal of geographical information science (IJGIS)*,
855 27(1):47–67.
- 856 Kyriakidis, P. (2004). A Geostatistical Framework For Area-To-Point Spatial
857 Interpolation. *Geographical Analysis*, 36.
- 858 Lam, N. S.-N. (1983). Spatial Interpolation Methods: A Review. *The Amer-*
859 *ican Cartographer*, 10(2):129–150.
- 860 Li, C., Lu, Z., Ma, T., and Zhu, X. (2009). A simple kriging method in-
861 corporating multiscale measurements in geochemical survey. *Journal of*
862 *Geochemical Exploration*, 101(2):147–154.
- 863 Lin, Y.-P., Cheng, B.-Y., Shyu, G.-S., and Chang, T.-K. (2010). Combining
864 a finite mixture distribution model with indicator kriging to delineate and
865 map the spatial patterns of soil heavy metal pollution in Chunghua County,
866 central Taiwan. *Environmental Pollution*, 158(1):235–244.
- 867 Moraga, P., Cramb, S. M., Mengersen, K. L., and Pagano, M. (2017). A
868 geostatistical model for combined analysis of point-level and area-level data
869 using INLA and SPDE. *Spatial Statistics*, 21:27–41.
- 870 Pedrycz, W. (2013). *Granular computing : analysis and design of intelligent*
871 *systems*. Industrial electronics series. Taylor & Francis, Boca Raton.
- 872 Pereira, O. J. R., Melfi, A. J., Montes, C. R., and Lucas, Y. (2018). Down-
873 scaling of ASTER Thermal Images Based on Geographically Weighted
874 Regression Kriging. *Remote Sensing*, 10(4):633.

- 875 Poggio, L. and Gimona, A. (2015). Downscaling and correction of regional cli-
876 mate models outputs with a hybrid geostatistical approach. *Spatial Statis-*
877 *tics*, 14:4–21.
- 878 Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian processes for ma-*
879 *chine learning*. Adaptive computation and machine learning. MIT Press,
880 Cambridge, Mass.
- 881 Rocas, M., García-González, A., Zlotnik, S., Larráyoz, X., and Díez, P.
882 (2021). Nonintrusive uncertainty quantification for automotive crash prob-
883 lems with VPS/Pamcrash. *Finite Elements in Analysis and Design*,
884 193:103556.
- 885 Roksvåg, T., Steinsland, I., and Engeland, K. (2021). A Two-Field Geosta-
886 tistical Model Combining Point and Areal Observations—A Case Study of
887 Annual Runoff Predictions in the Voss Area. *Journal of the Royal Statis-*
888 *tical Society Series C: Applied Statistics*, 70(4):934–960.
- 889 Schetelat, P., Lefort, L., and Delgado, N. (2020). Urban data imputation
890 using multi-output multi-class classification. *Building to Buildings: Urban*
891 *and Community Energy Modelling*.
- 892 Smith, B. J., Yan, J., and Cowles, M. K. (2008). Unified Geostatistical
893 Modeling for Data Fusion and Spatial Heteroskedasticity with R Package
894 ramps. *Journal of Statistical Software*, 25:1–21.
- 895 Truong, P. and Heuvelink, G. (2013). Bayesian Area-to-Point Kriging using
896 Expert Knowledge as Informative Priors. *International Journal of Applied*
897 *Earth Observation and Geoinformation*, 30:2291.
- 898 Wang, Q., Shi, W., and Atkinson, P. M. (2016). Area-to-point regression
899 kriging for pan-sharpening. *ISPRS Journal of Photogrammetry and Re-*
900 *mote Sensing*, 114:151–165.
- 901 Williams, C. and Rasmussen, C. (1996). Gaussian Processes for Regression.
902 In *Proceedings of the 1995 Conference*, Cambridge, Mass. The MIT Press.
- 903 Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R,*
904 *Second Edition*. CRC Press, Boca Raton.

- 905 Yoo, E.-H. and Kyriakidis, P. C. (2006). Area-to-point Kriging with
906 inequality-type data. *Journal of Geographical Systems*, 8(4):357–390.
- 907 Zadeh, L. A. (2005). Toward a generalized theory of uncertainty (GTU)—an
908 outline. *Information Sciences*, 172(1):1–40.
- 909 Zhang, X., Zuo, W., Zhao, S., Jiang, L., Chen, L., and Zhu, Y. (2018).
910 Uncertainty in Upscaling In Situ Soil Moisture Observations to Multiscale
911 Pixel Estimations with Kriging at the Field Level. *ISPRS International*
912 *Journal of Geo-Information*, 7(1).

913 **Appendix A. Proof of Proposition 2 (supplementary material)**

914 This proof employs a classical statistical approach to compute a Best
 915 Linear Unbiased Predictor using a family of observed random variables that
 916 are not necessarily Gaussian but have known first and second moments.

917 It is interesting for the understanding of the problem to give it a geomet-
 918 rical approach. Let us denote $F_i(g)$ the set of linear unbiased predictors of
 919 $Y_i(g)$ given an observation vector $\underline{\mathbf{Y}}$. With previous notations, it means that:

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

920 And similarly, we denote:

$$\begin{aligned} G_i(g) &:= \{ \alpha Y_i(g) : \alpha \in \mathbb{R} \} \\ F &:= \{ \boldsymbol{\alpha}^\top \underline{\mathbf{Y}} : \boldsymbol{\alpha} \in \mathbb{R}^n \} \quad (\text{the feature space generated by observations}) \\ F_0 &:= \{ \boldsymbol{\alpha}^\top \underline{\mathbf{Y}} : \boldsymbol{\alpha}^\top \underline{\boldsymbol{\mu}} = 0 \} \\ H &:= F \times G_i(g) \end{aligned}$$

One can note that F_0 is a subspace of F of dimension $\dim(F) - 1$. More-
 over $F_0 + F_i(g) = F_i(g)$, meaning that $F_i(g)$ is an affine subspace of F having
 F_0 for underlying vector space (see Figure A.9). But it also means that the
 sets of unbiased linear predictors for each output variable are parallel:

$$\forall i, j \in \{1, \dots, p\}, \forall g, g' \in \chi, F_i(g) \parallel F_j(g')$$

921 Now, given that we are minimizing the quadratic error between $Y_i(g)$ and
 922 $M_i(g)$, which can be seen as the distance between $Y_i(g)$ and $M_i(g)$ in H , the
 923 optimization process is geometrically a projection of $Y_i(g)$ on $F_i(g)$. This
 924 approach is illustrated in Figure A.9.

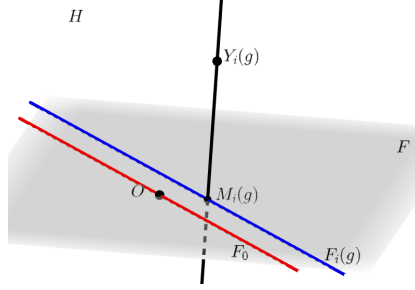


Figure A.9: Geometrical interpretation of the prediction process.

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

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

(i) If $\underline{\boldsymbol{\mu}} = (0, \dots, 0)^{\top}$ 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 ,

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

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

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

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

931 (ii) If $\underline{\boldsymbol{\mu}} \neq (0, \dots, 0)^{\top}$ then the condition for unbiasedness writes $\mu_i(g) =$
 932 $\alpha^{\top} \underline{\boldsymbol{\mu}}$ by linearity of expectation.

$v_i(g, \alpha)$ rewrites again:

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

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

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

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

Let us consider now the cross-errors:

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

934 Due to unbiasedness condition, it means that:

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

935 Which rewrites:

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

936 Note that equation (A.1) 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:

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

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

938 **Appendix B. Cross-errors and conditional covariances (supple-**
939 **mentary material)**

940 It is well known that the best predictor of $Y_i(g)$ is also the conditional
941 variable $\mathbb{E}[Y_i(g)|\underline{\mathbf{Y}}]$. However, this best predictor is not necessarily linear,
942 especially in non Gaussian cases. The following proposition proves that if the
943 Best Linear Unbiased Predictor is the best overall predictor then the error
944 covariances can also be seen as conditional covariances.

Proposition 3 (Cross-errors and conditional covariances). *Consider the as-*
sumption

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

Under assumption (A), cross errors for both Simple Mixture Kriging and Ordinary Mixture Kriging are:

$$c_{i,j}(g, g') = \mathbb{E}[\text{Cov}[Y_i(g), Y_j(g')|\underline{\mathbf{Y}}]]. \quad (\text{B.1})$$

Moreover, if $\text{Cov}[Y_i(g), Y_j(g')|\underline{\mathbf{Y}}]$ does not depend on $\underline{\mathbf{Y}}$, as it is the case for conditional Gaussian vectors, Equation (B.1) simplifies: $\mathbb{E}[\text{Cov}[Y_i(g), Y_j(g')|\underline{\mathbf{Y}}]] = \text{Cov}[Y_i(g), Y_j(g')|\underline{\mathbf{Y}}]$.

945 Assumption (A) holds for example when $\{\mathbf{Y}(x) : x \in \chi\}$ is a vector-
946 valued Gaussian random field and when each X_g is Dirac distributed.

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

950 • First, given the notations introduced in Appendix A, let $\delta \in F_0$ be a
951 non-zero vector and β a real number.

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

954 We have:

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

955 The minimum value of this polynomial expression is reached for:

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

956 Since the only optimal point is $M_i(g)$, $M_i^{\beta_0}(g) = M_i(g)$ and therefore
 957 $\beta_0 = 0$. As a consequence, as both $\mathbb{E}[\epsilon_i(g)] = 0$ and $\mathbb{E}[\delta] = 0$:

$$\forall \delta \in F_0, \forall i \in \{1, \dots, p\}, \forall g \in \chi, \quad \mathbb{E}[\epsilon_i(g) \delta] = \text{Cov}[\epsilon_i(g), \delta] = 0. \quad (\text{B.2})$$

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

963 • Now, let δ and δ' be any two vectors of F_0 . As a consequence of the
 964 previous result in Equation (B.2), we have:

$$\text{Cov}[\epsilon_i(g) + \delta, \epsilon_j(g') + \delta'] = c_{i,j}(g, g') + 0 + 0 + \text{Cov}[\delta, \delta'] \quad (\text{B.3})$$

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

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

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

$$M_i(g) = \mathbb{E}[Y_i(g) \mid \underline{\mathbf{Y}}] \text{ and } M_j(g') = \mathbb{E}[Y_j(g') \mid \underline{\mathbf{Y}}],$$

966 therefore

$$\begin{aligned} \mathbb{E}[\epsilon_i(g) \mid \underline{\mathbf{Y}}] &= \mathbb{E}[\epsilon_j(g') \mid \underline{\mathbf{Y}}] = 0 \\ \text{and } \text{Cov}[\epsilon_i(g) + \delta, \epsilon_j(g') + \delta'] &= \mathbb{E}[\text{Cov}[Y_i(g), Y_j(g') \mid \underline{\mathbf{Y}}]] + \text{Cov}[\delta, \delta'] \end{aligned} \quad (\text{B.4})$$

967 Identifying the equations (B.3) and (B.4), we get the expected result.

968 □

969	Contents	
970	1 Introduction	2
971	1.1 Classifying the EPC prediction problem in research	2
972	1.2 The limits of systematic averaging for spatial interpolation . .	5
973	1.3 Beyond systematic averaging	7
974	2 Prediction model	9
975	2.1 Data model	9
976	2.2 Mean and covariances of output variables	12
977	2.3 Best unbiased linear predictor	14
978	2.4 Particular cases	17
979	3 Illustration	19
980	3.1 Unidimensional case: rounded inputs	19
981	3.2 Unidimensional case: grains of varying size	22
982	3.3 Energy Performance Certificate (EPC) prediction	25
983	4 Discussion and conclusion	32
984	Bibliography	36
985	Appendix A Proof of Proposition 2 (supplementary mate-	
986	rial)	i
987	Appendix B Cross-errors and conditional covariances (sup-	
988	plementary material)	v
989	Contents	vii