Fast Update of Conditional Simulation Ensembles

Clément Chevalier, Xavier Emery, David Ginsbourger

Fast Update of Conditional Simulation Ensembles

Clément Chevalier, Xavier Emery, and David Ginsbourger

Abstract Gaussian random fields (GRF) conditional simulation is a key ingredient in many spatial statistics problems for computing Monte-Carlo estimators and quantifying uncertainties on non-linear functionals of GRFs conditional on data. Conditional simulations are known to often be computer intensive, especially when appealing to matrix decomposition approaches with a large number of simulation points. Here we study the settings where conditioning observations are assimilated batch-sequentially, i.e. one point or batch of points at each stage. Assuming that conditional simulations have been performed at a previous stage, we aim at taking advantage of already available sample paths and by-products in order to produce updated conditional simulations at minimal cost. We provide explicit formulas allowing to update an ensemble of sample paths conditioned on $n \geq 0$ observations to an ensemble conditioned on $n+q$ observations, for arbitrary $q \geq 1$. Compared to direct approaches, the proposed formulas prove to substantially reduce computational complexity. Moreover, these formulas enable explicitly exhibiting how the q "new" observations are updating the "old" sample paths. Detailed complexity calculations highlighting the benefits of our approach with respect to state-of-the-art algorithms are provided and are complemented by numerical experiments.

Keywords Gaussian random fields · kriging residual algorithm · batch-sequential strategies · kriging update equations

C. Chevalier
IMSV, University of Bern, Alpeneggstrasse 22, 3012 Bern, Switzerland
e-mail: clement.chevalier@stat.unibe.ch

X. Emery
Department of Mining Engineering and Advanced Mining Technology Center, University of Chile, Avenida Tupper 2069, Santiago, Chile
e-mail: xemery@ing.uchile.cl

D. Ginsbourger (✉)
IMSV, University of Bern, Alpeneggstrasse 22, 3012 Bern, Switzerland
e-mail: david.ginsbourger@stat.unibe.ch

1 Introduction

Throughout the paper, $Z = (Z(\mathbf{x}))_{\mathbf{x}\in\mathbb{X}}$ is a random field defined on a probability space $(\Omega, \mathscr{B}, P)$, with index $\mathbf{x}$ lying in a bounded set $\mathbb{X} \subset \mathbb{R}^d$ ($d \geq 1$). The random field Z is assumed to be evaluated sequentially, first at n points $\mathbf{X}_n := (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{X}^n$ ($n \geq 0$), and then at q additional points $\mathbf{X}_q := (\mathbf{x}_{n+1}, \ldots, \mathbf{x}_{n+q}) \in \mathbb{X}^q$ ($q \geq 1$). A crucial assumption here concerning the random field Z is that its distribution at stage n be Gaussian. This includes of course the case when Z is a Gaussian Random Field (GRF), but also the case of intrinsic random fields with Gaussian generalized increments (Matheron, 1973) and the Bayesian settings (Omre and Halvorsen, 1989; Handcock and Stein, 1993) where Z is Gaussian conditionally on some linear trend parameters with improper uniform distribution, and $n \geq 1$ pointwise evaluations of Z are already available at stage n.

Assuming that M simulations of Z have been performed at stage n, we investigate procedures to "update" them when a vector of new observations, $Z(\mathbf{X}_q)$, is assimilated. More precisely, the goal is to get a fast algorithm that generates M sample paths, rigorously drawn from the distribution of Z conditional on all $n+q$ evaluations, by recycling previous simulations and calculations as much as possible.

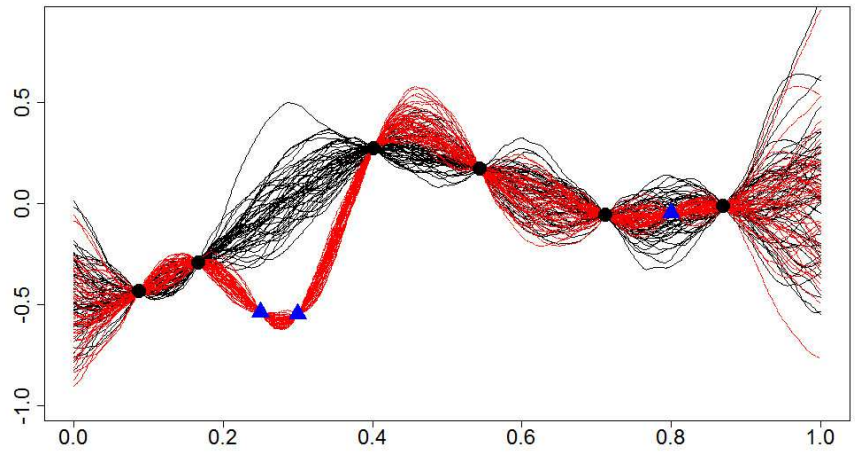

Fig. 1 GRF simulations conditioned on $n = 6$ observations (black curves) and $n+q = 9$ observations (red curves). The black circles stand for $n = 6$ initial observations and the blue triangles represent $q = 3$ additional observations.

The main contribution of this paper is illustrated on Fig. 1. An ensemble of 50 simulations of a GRF Z are performed conditionally on $n = 6$ observations (black curves). Our fast simulation update procedure (the "FOXY algorithm") is then ap-

plied to this ensemble in order to condition it on $q = 3$ additional observations at points $\mathbf{X}_q = (0.25, 0.3, 0.8)$ (red curves). Z is here assumed centred, possessing a non-conditional Matérn covariance with regularity parameter $\nu = 3/2$.

The motivations for developing such a procedure come from problems in different application fields. While GRF conditional simulations constitute a standard and important topic in the literature of geostatistics (Chilès and Delfiner, 2012) with a variety of applications in geosciences and natural resources characterization (Delhomme, 1979; Chilès and Allard, 2005; Deutsch, 2002; Dimitrakopoulos, 2011; Journel and Kyriakidis, 2004), they have been increasingly used in engineering and related areas, where Gaussian random field models have been used as prior distributions on expensive-to-evaluation functions (Hoshiya, 1995; Santner et al., 2003; Villemonteix et al., 2009; Roustant et al., 2012; Binois et al., 2014).

As conditional simulation methods are known to be generally computer intensive, an important challenge is the reduction of computation time through efficient algorithms. We consider the practically often encountered setting where observations are assimilated sequentially. In such case, when both new observations and previously simulated sample paths are available, it is obviously tempting to take advantage of the latter for obtaining sample paths conditioned on all observations without having to restart everything from scratch. A well-known algorithm which may apply to our settings is the residual substitution Z approach (Chilès and Delfiner, 2012; de Fouquet, 1994), also called conditioning kriging, or kriging residual algorithm. This method is used to post-process simulations obtained by using algorithms such as the circulant-embedding, spectral or turning bands (Chilès and Delfiner, 2012; Emery and Lantuéjoul, 2006). In this case, one only needs to construct non-conditional simulations and the assimilation of data is achieved through the conditioning kriging step (Hernández and Emery, 2009).

In the present paper, we provide efficient formulas allowing a fast "update" of GRF sample paths, together with a detailed algorithm, FOXY, the complexity of which is studied in detail to justify the improvement with respect to the kriging residual algorithm. The new formulas have the advantage of analytically exhibiting the dependence between the updated GRF sample path and the newly assimilated observations $\mathbf{Z}_{\text{new}} := (Z(\mathbf{x}_{n+1}), \ldots, Z(\mathbf{x}_{n+q})) = Z(\mathbf{X}_q)$. One of the key ingredient to obtain the formulas and set up the algorithm happens to be the *batch-sequential kriging update formulas* of Emery (2009); Chevalier et al. (2014), as detailed next. The paper is organized as follows: in Sect. 2 we revisit two already well-established approaches, namely the kriging residual algorithm and the kriging update formulas, and we then build upon them to derive our proposed *fast update of conditional simulation ensemble* ("FOXY") formula. Related algorithms and their complexity are presented in Sect. 3. Subsequent numerical experiments illustrating the efficiency of the FOXY algorithm are finally given in Sect. 4. For brevity and self-containedness, basics of kriging and more detailed versions of the algorithms are given in appendix.

2 Theory: from residual kriging to the FOXY algorithm

This section gives the main result of the paper. Sects. 2.1 and 2.2 detail two crucial ingredients which are used to obtain a new update formula in Sect. 2.3.

2.1 Kriging residual algorithm: old and new

In the simple kriging settings, the *kriging residual algorithm* or *residual substitution* approach (Chilès and Delfiner, 2012) is known to provide a simple and efficient way to produce simulations of a Gaussian Random Field Z conditional on observations at $\mathbf{X}_q$ ($q \geq 1$) relying both on non-conditional simulations of Z and some replicates of it, and on simple kriging means of Z and these replicates given their respective values at $\mathbf{X}_q$. Denoting by $z(\mathbf{x}_i)$ $(1 \leq i \leq q)$ the values of Z observed at $\mathbf{X}_q$, by $z^{(i)}$ $(1 \leq i \leq M)$ the non-conditional realizations of Z, and by $\mathbf{E}_p := \{\mathbf{e}_1, \ldots, \mathbf{e}_p\} \subset \mathbb{X}$ a considered set of simulation points (now assumed to be a finite subset of $\mathbb{X}$, for simplicity) the procedure consists of the following Algorithm 1.

Algorithm 1 Standard kriging residual algorithm

Require: The distribution of the Gaussian random field Z
Require: Evaluation points $\mathbf{X}_q = \{\mathbf{x}_1, \ldots, \mathbf{x}_q\} \subset \mathbb{X}$
Require: Evaluation results $z(\mathbf{x}_i)$ $(1 \leq i \leq q)$
Require: Simulation points $\mathbf{E}_p = \{\mathbf{e}_1, \ldots, \mathbf{e}_p\} \subset \mathbb{X}$

Step 1. Simulate M replicates of Z at $\mathbf{X}_q \cup \{\mathbf{e}_1, \ldots, \mathbf{e}_p\}$, denoted $z^{(i)}$ $(1 \leq i \leq M)$

Step 2. Calculate the kriging mean function m of Z knowing $z(\mathbf{x}_i)$ $(1 \leq i \leq q)$, and evaluate it at the simulation points $\mathbf{E}_p$, delivering the vector $m(\mathbf{E}_p)$.

Step 3.
for $j = 1 \to M$ **do**

- Calculate the kriging mean function $m^{(i)}$ of Z knowing $z^{(i)}(\mathbf{X}_q)$, and evaluate it at the simulation points $\mathbf{E}_p$, delivering the vector $m^{(i)}(\mathbf{E}_p)$. The kriging weights are the same as that calculated at Step 2.
- Set $r^{(i)}(\mathbf{E}_p) = z^{(i)}(\mathbf{E}_p) - m^{(i)}(\mathbf{E}_p)$
- Set $\mathbf{z}^{\star(\mathbf{i})} = m(\mathbf{E}_p) + r^{(i)}(\mathbf{E}_p)$

end for
Return $\mathbf{z}^{\star(\mathbf{1})}, \ldots, \mathbf{z}^{\star(\mathbf{M})}$ as conditional simulations of Z at $\mathbf{E}_p$ knowing $Z(\mathbf{X}_q) = z(\mathbf{X}_q)$.

While this algorithm seems to be common knowledge in geosciences, it is hard to find more than a few lines of expedited mathematical justifications of it across the geostatistics literature (Chilès and Delfiner, 2012; de Fouquet, 1994). Actually, this procedure turns out to be valid in more generality than the framework in which it is usually presented, as we prove below.

Proposition 1 (A generalization of the kriging residual algorithm) *Let $q \geq 1$, $\mathbf{X}_q = \{\mathbf{x}_1, \ldots, \mathbf{x}_q\} \subset \mathbb{X}$, Z be a random field such that $Z|Z(\mathbf{X}_q)$ is Gaussian with condi-*

tional expectation $M_q := \mathbb{E}(Z|Z(\mathbf{X}_q))$*, and* $Z^{(1)},\ldots,Z^{(M)}$ $(M \geq 1)$ *be independent replicates of* Z *with* $M_q^{(i)} := \mathbb{E}(Z^{(i)}|Z^{(i)}(\mathbf{X}_q))$ $(1 \leq i \leq M)$*. Then, the random fields*

$$Z^{\star(i)} = M_q + Z^{(i)} - M_q^{(i)} \quad (1 \leq i \leq M)$$

are equal to Z *in distribution, so that conditional on the event* $Z(\mathbf{X}_q) = \mathbf{z}_q$ *(for arbitrary* $\mathbf{z}_q \in \mathbb{R}^q$*) and denoting by* m_q *the corresponding realization of* M_q*,*

$$(Z|Z(\mathbf{X}_q) = \mathbf{z}_q) \overset{\mathscr{L}}{=} m_q + Z^{(i)} - M_q^{(i)} \tag{1}$$

holds for all $1 \leq i \leq M$*. Furthermore, the random fields* $m_q + Z^{(i)} - M_q^{(i)}$ *are stochastically independent.*

Proof. By definition of the conditional expectation, we have the decomposition

$$Z = M_q + R_q, \tag{2}$$

where $M_q = \mathbb{E}(Z|Z(\mathbf{X}_q))$ depends on Z only through its values at $\mathbf{X}_q$ (technically, M_q is $\sigma(Z(\mathbf{X}_q))$-measurable) and $R_q := (Z - M_q)$ is independent of $Z(\mathbf{X}_q)$. The same straightforwardly applies to the replicates $Z^{(1)},\ldots,Z^{(M)}$, and we use the similar notation $Z^{(i)} = M_q^{(i)} + R_q^{(i)}$, where the $M_q^{(i)}$'s are $\sigma(Z^{(i)}(\mathbf{X}_q))$-measurable and the $R_q^{(i)}$'s are respectively independent of $Z^{(i)}(\mathbf{X}_q)$ $(1 \leq i \leq M)$. Defining

$$Z^{\star(i)} = M_q + Z^{(i)} - M_q^{(i)} = M_q + R_q^{(i)} \quad (1 \leq i \leq M)$$

and using the fact that $R_q^{(i)}$ and R_q have same distribution and are both independent of M_q, we easily obtain that $Z^{\star(i)} \overset{\mathscr{L}}{=} Z$. Besides, by independence between the residuals and the values of the respective random fields at points $\mathbf{X}_q$, we get

$$(Z^{\star(i)}|Z(\mathbf{X}_q) = \mathbf{z}_q) \overset{\mathscr{L}}{=} (m_q + R_q^{(i)}) \overset{\mathscr{L}}{=} (m_q + R_q) \overset{\mathscr{L}}{=} (Z|Z(\mathbf{X}_q) = \mathbf{z}_q) \tag{3}$$

Finally, the $Z^{\star(i)}$'s are indeed independent conditionally on $Z(\mathbf{X}_q)$ by independence of the $R_q^{(i)}$'s. □

An example of application of Proposition 1 in a non-standard set up is given in Fig. 2. A Universal Kriging model is assumed, in which Z has already been evaluated at n (not represented) points prior to the evaluation at the q conditioning points, so that Z's distribution at stage n is indeed Gaussian but with non-stationary mean and covariance kernel given by the Universal Kriging equations (recalled in appendix).

From that perspective, the conditional expectation of Z when $q = 3$ new observations are available can be obtained using simple kriging with the previous non-stationary mean and covariance kernel. Let us stress here that computing simple kriging means with a non-conditional covariance function being the universal kriging covariance function based on n past observations is not necessarily sensible from

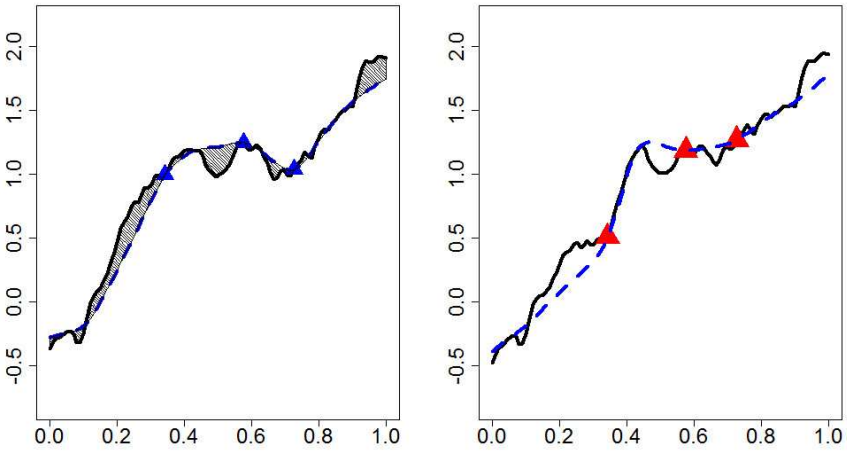

Fig. 2 Left: kriging residual $r^{(i)} = z^{(i)} - m_q^{(i)}$ obtained by non-conditional simulation of a replicate $Z^{(i)}$ of a non-stationary GRF Z (black solid line) and its simple kriging mean (blue dashed line) based on $q = 3$ observations (blue triangles) at a design $\mathbf{X}_q$. Right: conditional simulation of Z (solid black line) obtained by summing Z's simple kriging mean (blue dashed line) based on its values at the same design $\mathbf{X}_q$ (red triangles) and the kriging residual $r^{(i)}$ simulated on the left graph.

a computational point of view, but that having this particular approach in mind will facilitate understanding forthcoming ideas. It will appear in the next section on update, though, that such non-stationary simple kriging mean actually coincides with the usual Universal Kriging mean relying on all $n+q$ observations.

2.2 Kriging update framework

Keeping in mind the overall set up of a random field whose distribution at stage n is Gaussian with mean M_n (the realization of which is denoted m_n) and covariance k_n, let us now focus on the situation where a batch of evaluations at $q > 0$ additional points $\mathbf{X}_q := \{\mathbf{x}_{n+1}, \ldots, \mathbf{x}_{n+q}\}$ is assimilated.

In the literature, a lot of effort has been paid to obtain update formulas for the kriging predictors when observations are assimilated sequentially. Update formulas are meant to enable a fast computation of the kriging mean M_{n+q} and the kriging covariance function k_{n+q} in case M_n and k_n are already available. Barnes and Watson (1992) gave kriging update formulas for the kriging mean and covariance function in Simple Kriging settings, with $q = 1$. Gao et al. (1996) generalized these formulas, still with $q = 1$, to Universal Kriging settings. Finally Emery (2009) obtained Universal Kriging update formulas for the kriging mean for arbitrary $q \geq 1$ and

Chevalier et al. (2014) complemented them with update formulas for the kriging covariance function. These formulas, adapted to our settings and notations are recalled below:

$$M_{n+q}(\mathbf{x}) = M_n(\mathbf{x}) + \lambda_{n,q}(\mathbf{x})^\top (Z(\mathbf{X}_q) - M_n(\mathbf{X}_q)) \tag{4}$$

$$k_{n+q}(\mathbf{x},\mathbf{x}') = k_n(\mathbf{x},\mathbf{x}') - \lambda_{n,q}(\mathbf{x})^\top K_{n,q} \lambda_{n,q}(\mathbf{x}') \tag{5}$$

$$\lambda_{n,q}(\mathbf{x}) = K_{n,q}^{-1} k_n(\mathbf{x}, \mathbf{X}_q) \tag{6}$$

where $\lambda_{n,q}(\mathbf{x})$ is a vector of q kriging weights of responses at $\mathbf{X}_q$ for predicting at point $\mathbf{x}$ at time n, and $K_{n,q} := (k_n(\mathbf{x}_{n+i}, \mathbf{x}_{n+j}))_{1\leq i,j\leq q}$. These formulas enable significant computational savings as only q kriging weights $\lambda_{n,q}$ (and not $n+q$) need to be computed for obtaining updated kriging mean and covariance functions. In particular, a cumbersome $(n+q)\times(n+q)$ matrix inversion is avoided.

Note that the kriging update formulas (4),(5) corroborates the fact that in the previous example (Fig. 2), performing simple kriging using M_n and k_n as non-conditional covariance functions and q new observations actually gave the same result as if we had performed Universal Kriging based on the initial model, with $n+q$ observations.

Before going further and presenting our update formulas for conditional simulations, let us briefly present a random field decomposition that enables easily retrieving update formulas (and extending them without difficulty to our conditionally Gaussian framework), and that will be crucial for the main result.

Proposition 2 (Three-stage telescopic decomposition) *Let Z be a random field with Gaussian distribution conditionally on its observation at $\mathbf{X}_n$, and $\mathbf{X}_q$ be additional points. Then Z decomposes as*

$$Z = M_n + \lambda_{n,q}^\top R_n(\mathbf{X}_q) + R_{n+q}, \tag{7}$$

where $R_{n+q} := Z - \mathbb{E}(Z|Z(\mathbf{X}_n), Z(\mathbf{X}_q))$ is a centred Gaussian random field independent of $Z(\mathbf{X}_n)$ and $Z(\mathbf{X}_q)$, and $\lambda_{n,q}$ is defined in Eq. 6.

Proof. From the previous section on the residual kriging algorithm, we already know that $Z = M_n + R_n$ with $M_n = \mathbb{E}(Z|Z(\mathbf{X}_n))$, and where $R_n = Z - \mathbb{E}(Z|Z(\mathbf{X}_n))$ is a centred Gaussian random field independent of $Z(\mathbf{X}_n)$, with covariance k_n. Applying a similar decomposition to R_n with respect to the two batches of observations, $Z(\mathbf{X}_n)$ and $Z(\mathbf{X}_q)$, we obtain that

$$Z = M_n + \mathbb{E}(R_n|Z(\mathbf{X}_q), Z(\mathbf{X}_n)) + R_{n+q},$$

where $R_{n+q} := R_n - \mathbb{E}(R_n|Z(\mathbf{X}_q)) = Z - \mathbb{E}(Z|Z(\mathbf{X}_n), Z(\mathbf{X}_q))$ is a centred Gaussian random field, independent of $Z(\mathbf{X}_n)$ and $Z(\mathbf{X}_q)$ by fundamental property of the conditional expectation. There remains to notice that $\mathbb{E}(R_n|Z(\mathbf{X}_q), Z(\mathbf{X}_n)) = \mathbb{E}(R_n|R_n(\mathbf{X}_q), Z(\mathbf{X}_n)) = \mathbb{E}(R_n|R_n(\mathbf{X}_q))$ by independence between the field R_n and the random vector $Z(\mathbf{X}_n)$, and finally that $\mathbb{E}(R_n|R_n(\mathbf{X}_q)) = \lambda_{n,q}^\top R_n(\mathbf{X}_q)$ precisely because R_n is a centred Gaussian field with covariance kernel k_n. □

Let us mention as evocated earlier that taking conditional expectations on both sides (with respect to $Z(\mathbf{X}_n), Z(\mathbf{X}_q)$) in Eq. 7 directly delivers Eq. 4 while taking conditional covariances leads to Eq. 5 without much effort.

2.3 FOXY: Fast Update of Conditional Simulation Ensembles

We now have the ingredients to detail a new update formula which will serve as a basis to our method for fast updating ensembles of conditional simulations, referred to as the FOXY algorithm. Let us now state a central result of the paper, all notations being kept as in the previous sections unless precised otherwise.

Proposition 3 (Conditional simulation update formula) *Let $Z^{(1)},\ldots,Z^{(M)}$ be independent replicates of $Z|Z(\mathbf{X}_n)$, i.e., simulations of Z conditioned on the n observations $Z(\mathbf{X}_n)$. Then, the random fields*

$$Z^{\star(i)} := Z^{(i)} + \lambda_{n,q}^{\top}(Z(\mathbf{X}_q) - Z^{(i)}(\mathbf{X}_q)) \quad (i \in \{1,\ldots,M\}) \tag{8}$$

have the same conditional distribution as Z conditioned on the $n+q$ observations $Z(\mathbf{X}_n), Z(\mathbf{X}_q)$, i.e. for any conditioning values $\mathbf{z}_n \in \mathbb{R}^n$, $\mathbf{z}_q \in \mathbb{R}^q$,

$$\mathscr{L}(Z|Z(\mathbf{X}_n) = \mathbf{z}_n, Z(\mathbf{X}_q) = \mathbf{z}_q) = \mathscr{L}(Z^{(i)} + \lambda_{n,q}^{\top}(\mathbf{z}_q - Z^{(i)}(\mathbf{X}_q)) \tag{9}$$

Furthermore, the $Z^{(i)} + \lambda_{n,q}^{\top}(\mathbf{z}_q - Z^{(i)}(\mathbf{X}_q))$'s are stochastically independent.

Proof. Let $M_{n+q} := \mathbb{E}(Z|Z(\mathbf{X}_n), Z(\mathbf{X}_q))$ and $M_{n+q}^{(i)}$ be defined similarly for the i^{th} replicate $Z^{(i)}$. The equality in distribution of Z and $Z^{(i)}$ implies that:

$$Z - M_{n+q} \overset{\mathscr{L}}{=} Z^{(i)} - M_{n+q}^{(i)} \tag{10}$$

Now, an application of the kriging update formula (4) to both M_{n+q} and $M_{n+q}^{(i)}$, together with the identity $M_n = M_n^{(i)}$, yields

$$\begin{aligned} Z &\overset{\mathscr{L}}{=} M_n + \lambda_{n,q}^{\top}(Z(\mathbf{X}_q) - M_n(\mathbf{X}_q)) + (Z^{(i)} - M_n^{(i)}) - \lambda_{n,q}^{\top}(Z^{(i)}(\mathbf{X}_q) - M_n^{(i)}(\mathbf{X}_q)) \\ &\overset{\mathscr{L}}{=} Z^{(i)} + \lambda_{n,q}^{\top}(Z(\mathbf{X}_q) - Z^{(i)}(\mathbf{X}_q)) \end{aligned}$$

which completes the proof. □

The formulas given in Proposition 3 offer many advantages. In particular, Eq. (8) explicitly quantifies the effect of the newly assimilated observations $Z(\mathbf{X}_q)$ on simulated GRF paths. This can be used in our settings, where we have already simulated GRF sample paths $z^{(1)},\ldots,z^{(M)}$, from $Z^{(1)},\ldots,Z^{(M)}$, to quickly obtain sample paths $z^{\star(1)},\ldots,z^{\star(M)}$ sampled from $Z^{\star(1)},\ldots,Z^{\star(M)}$ of Eq. (8). In the next section, we will show that the proposed algorithm has a lower complexity than the classical kriging

residual algorithm, which would update the GRFs $Z^{(i)}, 1 \leq i \leq M$ by computing a kriging mean based on $n+q$ observations.

3 Complexity calculation

3.1 Kriging residual and FOXY algorithms

In this section, we investigate the computational complexity of simulating $M > 0$ sample paths of a GRF Z conditioned on $n+q$ observations at points $\mathbf{X}_n, \mathbf{X}_q$ in the case where M sample paths conditioned on $Z(\mathbf{X}_n)$ are already available. The result of the algorithms will be an ensemble of M (conditionally) independent realizations of a GRF with conditional mean and covariance, m_{n+q} and k_{n+q}, given by the kriging equations. We shall consider here the settings of UK (Eqs. (15),(13)), but the results and conclusions will be unchanged in the SK settings. Only the case $n > 0$ will be considered though, as, if $n = 0$, the two algorithms that we shall detail are exactly the same.

Here, we assume that all the simulations are performed at p simulation points $\mathbf{e}_1, \ldots, \mathbf{e}_p$ in $\mathbb{X}$. To simplify the complexity calculations we further assume that p, n are much larger than q. In UK, another variable is the number ℓ of basis functions for the trend (Appendix A). This variable is also assumed to be much smaller than p, n. To summarize, we assume that $p, n \gg q, \ell$.

Two major cases will be distinguished in the algorithms. First, an unfavorable case where the set of new observation points $\mathbf{X}_q$ is not included in the set of simulation points $\{\mathbf{e}_1, \ldots, \mathbf{e}_p\}$. Second, the favorable case where it is. Finally, two algorithms will be compared. The first one is a classical kriging residual algorithm, based on Eq. (1). This algorithm requires to compute a kriging mean M_{n+q} based on the real observations $Z(\mathbf{X}_n), Z(\mathbf{X}_q)$ and also to obtain a kriging residual, which involves another computation of a kriging mean based on the observations $Z(\mathbf{X}_n)$ and the "artificial" observations $Z^{(i)}(\mathbf{X}_q)$. These computations involve the calculation of $n+q$ kriging weights, which, in the particular case of UK, will be done using Eq. (11). The kriging weights are the same for the two kriging means that we compute. The second algorithm is the FOXY algorithm which is based on Proposition 3 and Eq. (8). FOXY has the advantage of requiring only the computation of $\lambda_{n,q}$, i.e. q kriging weights per simulation point. As we compute only q kriging weights, and not $n+q$, we shall show that FOXY brings a computational complexity reduction of $O(n/q)$ compared to the classical kriging residual algorithm. The complexities obtained will also be compared to the one of a third "benchmark" algorithm based on a decomposition (e.g., Cholesky) of the conditional covariance matrix at the simulation points (Davis, 1987). This last algorithm does not take advantage of any previous computations. The algorithms are summarized in the next two subsections and the details are given in Appendix B.

3.2 Preliminary step: computing $Z^{(i)}(\mathbf{X}_q)$ for all the M sample paths

An important detail that is relative to Eq. (1) and is even clearer in Eq. (8) is that the update of M GRF sample paths requires knowing, for each $1 \leq i \leq M$, the value of the sample path number i at the batch $\mathbf{X}_q$, which is denoted by $Z^{(i)}(\mathbf{X}_q)$. In Eq. (1) the knowledge of $Z^{(i)}(\mathbf{X}_q)$ is required, as the computation of the kriging residual involves a kriging mean based on q "artificial" observations $Z^{(i)}(\mathbf{X}_q)$. Thus, if the batch of q points $\mathbf{X}_q$ is not included in the set of p simulation points $\{\mathbf{e}_1,\ldots,\mathbf{e}_p\}$, the value of $Z^{(i)}(\mathbf{X}_q)$ needs to be simulated, conditionally on $n+p$ observations, and for all $1 \leq i \leq M$. This case is referred to as the "unfavorable case". In UK, it involves the computation of the kriging weights in Eq. (11) of the $n+p$ points $\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{e}_1,\ldots,\mathbf{e}_p$ for the prediction at each of the q points $\mathbf{X}_q$. In particular, the inversion of a $(n+p)\times(n+p)$ matrix in Eq. (11) is done with a cumbersome $O((n+p)^3)$ complexity. The computation of the other terms of Eq. (11) has a cost which is dominated by this term in our settings. Finally, once the kriging weights are computed the value of $Z^{(i)}(\mathbf{X}_q))$ is simulated for all i with a cost of $O(Mpq)$. Thus, in the described unfavorable case, both the classical kriging residual algorithm and the new algorithm based on Eq. (8) have a cost, referred to as the "preliminary cost", which is of $O((n+p)^3+Mpq)$. A detailed algorithm justifying this complexity is given in Appendix B. We shall see that the preliminary cost tend to dominate the other costs of the two algorithms so that the savings provided by the FOXY algorithm will be lower in the unfavorable case, i.e. if $\mathbf{X}_q$ is not included in the set of p simulation points.

3.3 Computing kriging weights and updating the sample paths

We now assume that for all $1 \leq i \leq M$ the quantity $Z^{(i)}(\mathbf{X}_q)$ is known. This is the case either if $\mathbf{X}_q$ is included in the set of simulation points $\{\mathbf{e}_1,\ldots,\mathbf{e}_p\}$ (favorable case), or if the preliminary cost of $O((n+p)^3+Mpq)$ has been paid (unfavorable case).

For the classical kriging residual algorithm, the computation of kriging means requires calculating $n+q$ kriging weights of $\mathbf{x}_1,\ldots,\mathbf{x}_{n+q}$ for the prediction at points $\mathbf{e}_1,\ldots,\mathbf{e}_p$. This will be done using Eq. (11). The inversion of the covariance matrix K at $n+q$ points in Eq. (11) has a $O((n+q)^3)$ complexity. The p matrix-vector products $K^{-1}\mathbf{k}(\mathbf{x})$, where $\mathbf{x}$ takes all the values $\mathbf{e}_1,\ldots,\mathbf{e}_p$, has a $O(p(n+q)^2)$ cost. The computation of the other terms of Eq. (11) involves other complexities that are all dominated by the $O(p(n+q)^2)$ cost. We thus end-up with a total cost of $O((n+q)^3+p(n+q)^2)$. It is important to note that the $O((n+q)^3)$ complexity to invert K can be reduced to $O(qn^2)$ using matrix block-inversion formulas based on the Schur complement in the realistic case where the covariance matrix computed in the n (and not $n+q$) points $\mathbf{x}_1,\ldots,\mathbf{x}_n$ has already been inverted before, or decom-

posed with a standard decomposition. We shall assume that this is the case here, so that our only dominating term in the complexity if now of $O(p(n+q)^2)$.

The FOXY algorithm based on Proposition 3 and Eq. (1) requires the computation of the q kriging weights of $\mathbf{x}_{n+1},\ldots,\mathbf{x}_{n+q}$ for the prediction at points $\mathbf{e}_1,\ldots,\mathbf{e}_p$. These kriging weights are equal to $K_{n,q}^{-1}k_n(\mathbf{x},\mathbf{X}_q)$, as indicated by Eq. (6)), where $\mathbf{x}$ takes the values $\mathbf{e}_1,\ldots\mathbf{e}_p$, k_n is the kriging covariance defined in Eq. (13) and $K_{n,q}=k_n(\mathbf{X}_q,\mathbf{X}_q)$ is the $q\times q$ kriging covariance matrix at $\mathbf{X}_q$ based on n observations. The computation of all these kriging covariances can be performed using Eq. (13). In that case, the cost to invert the matrix K is of $O(n^3)$ but, here, we shall assume that this inverse has already been computed. The remaining matrix vector multiplications are performed at a cost of mainly $O((\ell+q)(n^2+pn))$ (see, Algorithm 4 for complete details). This final cost is lower than the $O(p(n+q)^2)$ cost obtained with the classical kriging residual algorithm, as we assumed that $n,p\gg\ell,q$. In particular, for $q>\ell$ and $p>n$, FOXY has a complexity of $O(pnq)$ vs. $O(pn^2)$ for the classical kriging residual algorithm.

Once the kriging weights are computed; the remaining cost in both the kriging residual algorithm and the new algorithm is of $O(Mpq)$. This is simply the cost to perform q multiplications for all the p simulation points $(\mathbf{e}_1,\ldots,\mathbf{e}_p)$ and all the M sample paths. The aggregated complexity of the two studied algorithms are summarized in Table 3.3, together with the complexity of a standard algorithm based on a decomposition of the covariance matrix, which does not take advantage of previous computations. We see that in the unfavorable case, FOXY brings a lower improvement with respect to the kriging residual algorithm, as the dominating terms in the complexity are $O((n+p)^3)$ and $O(Mpq)$ for both algorithms. On the other hand, FOXY is expected to be much faster than the classical kriging residual algorithm in the favorable case where $\mathbf{X}_q$ is a subset of the set of simulation points because the preliminary cost of $O((n+p)^3)$ is not paid.

Table 1 Theoretical complexity of a classical kriging residual algorithm (Eq. (1)), the FOXY algorithm based on Eq. (8) and an algorithm based on a decomposition of the covariance matrix at the observation points.

algorithm	preliminary calc.	kriging weights	simulation
Kriging residual algo.	$O((n+p)^3+Mpq)$	$O(p(n+q)^2)$	$O(Mpq)$
New algo.	$O((n+p)^3+Mpq)$	$O((\ell+q)(n^2+pn))$	$O(Mpq)$
Decomp.-based algo.			$O(p^3+Mp^2)$

4 Numerical experiments

In this section, we illustrate the theoretical complexity results on numerical experiments. We investigate the computation time to update M simulations of a GRF conditioned on n observations into simulations conditioned on $n+q$ observations. The simulations are performed at p simulation points. Here the simulated fields are GRFs in two dimensions. The function used to condition the values of the GRFs is the Branin-Hoo function (Jones et al., 1998), which is here defined on $\mathbb{X} = [0,1]^2$. The type of kriging considered here is the so-called Ordinary Kriging (OK) where the non-conditional mean function of the GRF Z is simply an unknown constant. With our notations, this corresponds to the case of a single basis function ($\ell = 1$). As the computation times are not very sensitive to ℓ (at least if ℓ does not take large values), only the sensitivity of the computation time to M, n, p, q will be investigated.

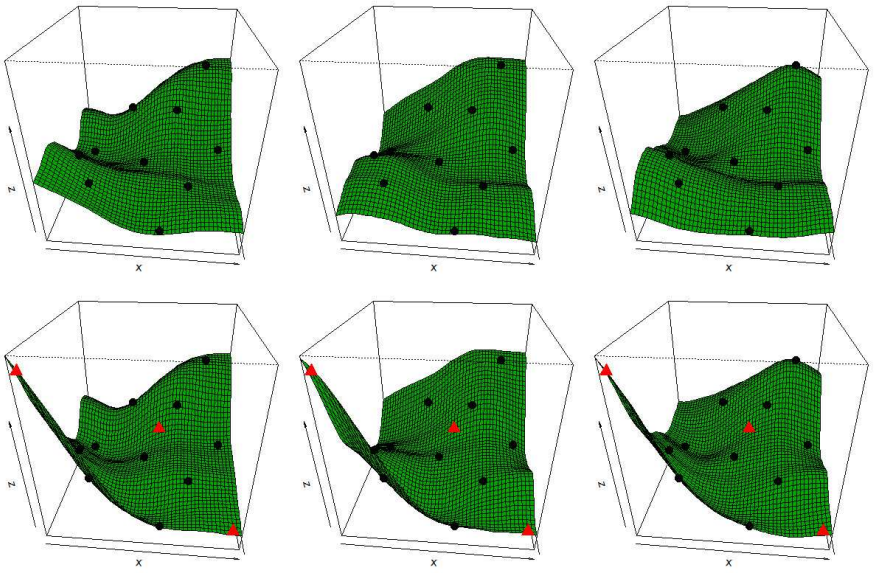

Fig. 3 Top: three realizations of a GRF in two dimension conditioned on $n = 10$ observations (black points). Bottom: update of these three realizations when $q = 3$ new observations (red triangles) are assimilated. The non-conditional covariance function of the random field is a Matérn covariance with parameter $\nu = 5/2$.

An example of GRFs in two dimensions, simulated on a grid of $50 \times 50 = 2500$ points is provided on Fig. 3. The three GRFs at the top are conditioned on $n = 10$ observations and each GRF is updated (using FOXY, here, see bottom plots) to be conditioned on $10+3$ observations. For our experiments, a set of values needs to be chosen for (M, n, p, q). The chosen set is the following:

$$(M,n,p,q) \in \mathbb{D}_M \times \mathbb{D}_n \times \mathbb{D}_p \times \mathbb{D}_q,$$

where $\mathbb{D}_M = \{1,100,1000,10000,20000,30000\}$, $\mathbb{D}_n = \{10,100,500,1000\}$, $\mathbb{D}_p = \{100,500,2000\}$ and $\mathbb{D}_q = \{1,10\}$. The set $\mathbb{D}_M$ covers cases where we update very few conditional simulations, so that the computation time will be dominated by the preliminary costs or by the time to compute kriging weights, and also cases where we update a large set of conditional simulations. Regarding $\mathbb{D}_n$, we will test cases where very few (i.e., 10) conditioning observations are available and also cases where we have up to 1000 conditioning observations. We chose not to go beyond a value of 1000 as standard uses of kriging are rarely done when the number of observations is larger because of an expensive $n \times n$ matrix inversion. For the same reason the values of p do not go beyond 2000. We recall that in the "unfavourable" case, a $(n+p) \times (n+p)$ matrix needs to be inverted. We may note however that, in the favorable case it is perfectly possible to use large p, or even infinite p. Finally, only two small values of q were considered as the change of computation time when q grows is rather simple, at least if q remains low. For all the possible triplets (n,p,q) we choose to plot the computation time as a function of M. The experiments were performed in both the unfavorable and favorable cases on a laptop with a 2.27 Ghz cpu and 3.7 Gb of RAM.

As detailed in the last section, the total computation time of the tested algorithms is the sum of the computation times of three different steps:

1. The preliminary cost (unfavorable case). Here both algorithms perform a $(n+p) \times (n+p)$ matrix inversion, which adds a fixed cost that does not depend on M.
2. The cost to compute kriging weights. In this step the FOXY algorithm computes q kriging weights while the kriging residual algorithm computes $n+q$ weights. This cost is also a fixed cost that does not depend on M. This step is where FOXY might be much faster than the classical algorithm.
3. The cost to update the simulations once the weights are computed. Both algorithm have an $O(Mpq)$ cost for this step. Hence, when q is large, the computation time is expected to grow faster with M.

Figure 4 detail our results in the favorable cases while Fig. 5 show the unfavorable case. Computation times are given in seconds. It is important to note that, for a given value of (n,p,q), the *difference* between the computations times of the two algorithms is the same in the favorable and unfavorable case. However, the *ratio* between these two computation times is not constant as the preliminary costs can be important. This explains why even if the time difference is unchanged, the computation times plotted on a log-scale seem to be closer for the two algorithms in the unfavorable case. The following conclusions can be drawn from Figs 4, 5. First, when n is very low, the time to compute kriging weights becomes negligible, which explains why the two algorithms have the same performances. In that case, the computation time is dominated by the $O(Mpq)$ complexity to update the M conditional simulations and, in the unfavorable case, by the preliminary costs (specially when p

is large).
When $n = 100$, the computation time to obtain the kriging weights is not negligible anymore. In the favorable case, the computation time is reduced with FOXY by a factor 2 for moderate (i.e., less than 1000) M. For larger M the performances tend to be similar as the $O(Mpq)$ complexity dominates again. In the unfavorable case, the computation time ratio between the two algorithms is lower than 2 because of the preliminary costs paid in both algorithms.
Finally, when $n = 500$ or 1000 the gap between the two algorithms gets larger. This is due to the time to compute the kriging weights that is $O(n/q)$ faster with FOXY. For $n = 1000$ FOXY can be up to 25 times faster in the favorable case and for moderate M. For large M, this factor decrease to approximately 10. In the unfavorable case, this factor is lower because of the preliminary costs. The higher these costs are, the lower the reduction factor is. For low p, FOXY is approximately 3 times faster, and only 2 times faster for large p.
The choice of q mainly influences the cost of the last step of both algorithms, which has a $O(Mpq)$ complexity. q is thus driving the slope of the curves presented on Figs. 4, 5. It is also interesting to note that, in absolute value, the computation times are generally low. With a very standard laptop, the update of $30,000$ simulations on 2000 points conditionally on 1000 observations takes "only" 1 second with FOXY in the favorable case, and 10 seconds in the unfavorable case. Performing the simulation "from scratch" using a Cholesky decomposition takes approximately 20 minutes, here.

5 Conclusion

In this paper, we provide efficient formulas allowing to quickly update ensembles of simulations of conditionally Gaussian random fields. Simulated paths, which are initially conditioned on $n \geq 0$ observations, are "twisted" in order to be conditioned on $n+q$ observations, for arbitrary $q \geq 1$. The formulas lead to a fast update algorithm which has been implemented in R and proves to offer substantial computational savings, especially when the number of conditioning observations, n is large. In addition, the formulas have the advantage of explicitly quantifying the effect of the q newly assimilated observations on the sample paths.

A limitation of the formulas though is that they apply only in the case where the covariance parameters of the non-conditional covariance function of the considered random field are assumed known. In the typical settings where the covariance parameters are re-estimated when new observations are assimilated, the formulas cannot be straightforwardly applied. Also, one may investigate to what extent a recursive use of the formulas may result in numerical instabilities.

The fast update approaches presented in this paper can be applied to efficiently compute Monte-Carlo estimates based GRF simulations in the case where the observations are assimilated sequentially. Example of potential applications include

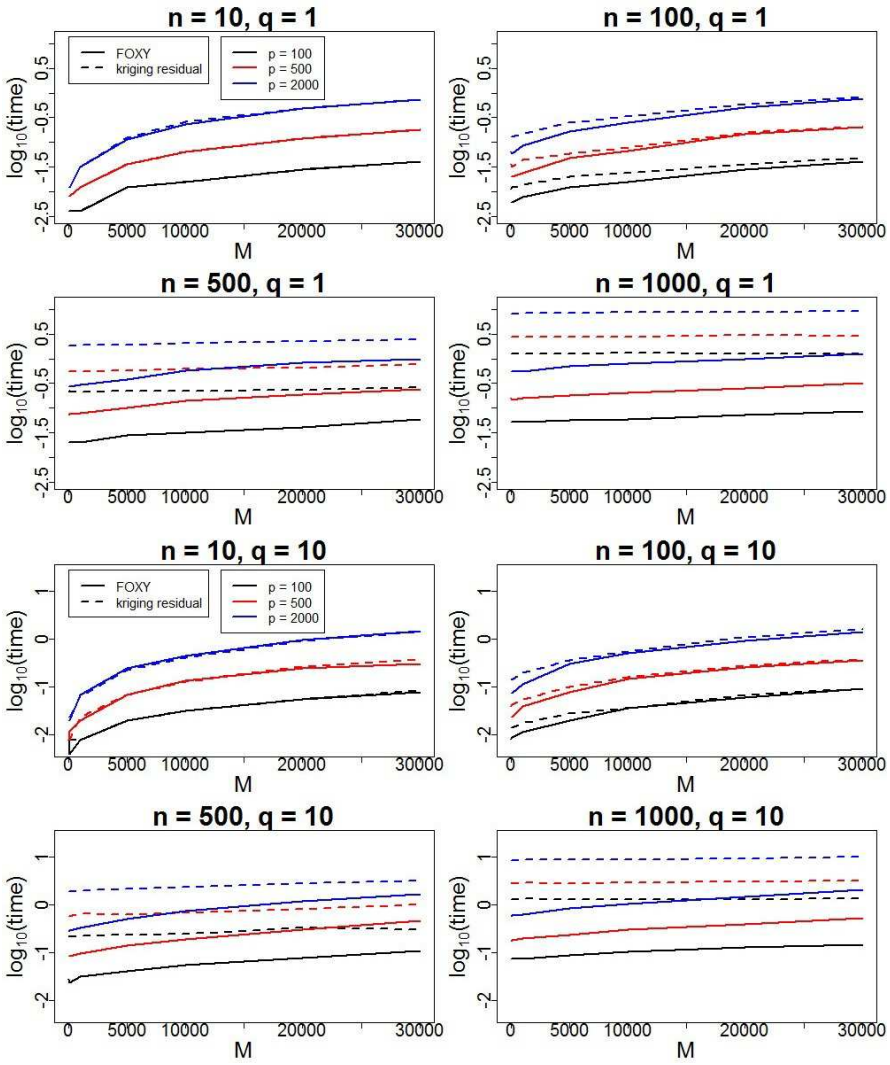

Fig. 4 Computation times in the favorable case.

estimating the Shanon entropy of the maximizer of a conditioned GRF, a crucial step in a recently proposed Bayesian global optimization algorithm (Villemonteix et al., 2009). In the same vein, updated simulations have been recently used in the framework of the "robust inversion" problem studied in Chevalier (2013). In any case, when relying on an ensemble conditional simulations, one has to keep in mind that potential biases in Monte-Carlo estimates due to sample finiteness may propagate along consecutive stages. When using such approach, and especially when it

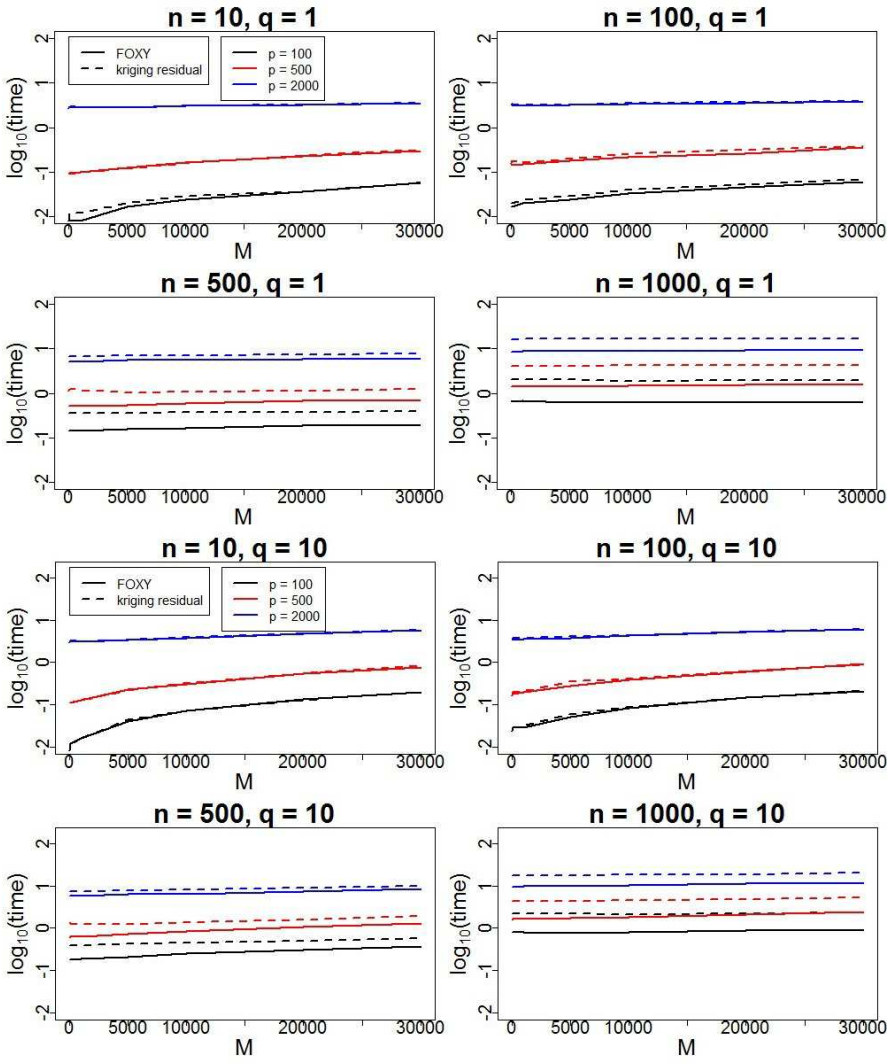

Fig. 5 Computation times in the unfavorable case.

comes to uncertainty quantification purposes, the sample size and the procedure for generating the initial set of simulations should hence be carefully chosen.

Acknowledgements Part of this work has been conducted within the frame of the ReDice Consortium, gathering industrial (CEA, EDF, IFPEN, IRSN, Renault) and academic (Ecole des Mines de Saint-Etienne, INRIA, and the University of Bern) partners around advanced methods for Computer Experiments. Clément Chevalier also gratefully acknowledges support from the French Nuclear Safety Institute (IRSN) and warmly thanks Prof. Julien Bect for fruitful discussions on GRF

simulation. David Ginsbourger acknowledges support from the IMSV and from the Swiss National Science Foundation (ENSEMBLE project, Sinergia grand No. CRSI22-132249/1).

References

Barnes, R. J. and Watson, A. (1992). Efficient updating of kriging estimates and variances. *Mathematical geology*, 24(1):129–133.

Binois, M., Ginsbourger, D., and Roustant, O. (2014). Quantifying uncertainty on pareto fronts with gaussian process conditional simulations. Technical report, ReDICE Consortium.

Chevalier, C. (2013). *Fast Uncertainty reduction strategies relying on Gaussian process models*. PhD thesis, University of Bern.

Chevalier, C., Ginsbourger, D., and Emery, X. (2014). Corrected kriging update formulae for batch-sequential data assimilation. In *Mathematics of Planet Earth*, pages 119–122. Springer.

Chilès, J. and Allard, D. (2005). Stochastic simulation of soil variations. In Grunwald, S., editor, *Environmental Soil-Landscape Modeling: Geographic Information Technologies and Pedometrics*, pages 289–321. CRC Press.

Chilès, J.-P. and Delfiner, P. (2012). *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York, second edition edition.

Davis, M. W. (1987). Production of conditional simulations via the lu triangular decomposition of the covariance matrix. *Mathematical Geology*, 19(2):91–98.

de Fouquet, C. (1994). Reminders on the conditioning kriging. In Armstrong, M. and Dowd, P., editors, *Geostatistical Simulations*, pages 131–145. Kluwer Academic Publishers.

Delhomme, J. (1979). Spatial variability and uncertainty in groundwater flow parameters: a geostatistical approach. *Water Resources Research*, 15(2):269–280.

Deutsch, C. (2002). *Geostatistical Reservoir Modeling*. Oxford University Press, New York.

Dimitrakopoulos, R. (2011). *Advances in Ore Body Modelling and Strategic Mine Planning*. Australasian Institute of Mining and Metallurgy, Melbourne.

Emery, X. (2009). The kriging update equations and their application to the selection of neighboring data. *Computational Geosciences*, 13(1):211–219.

Emery, X. and Lantuéjoul, C. (2006). Tbsim: A computer program for conditional simulation of three-dimensional gaussian random fields via the turning bands method. *Computers & Geosciences*, 32(10):1615–1628.

Gao, H., Wang, J., and Zhao, P. (1996). The updated kriging variance and optimal sample design. *Mathematical Geology*, 28:295–313.

Handcock, M. S. and Stein, M. L. (1993). A bayesian analysis of kriging. *Technometrics*, 35(4):403–410.

Hernández, J. and Emery, X. (2009). A geostatistical approach to optimize sampling designs for local forest inventories. *Canadian Journal of Forest Research*, 39:1465–1474.

Hoshiya, M. (1995). Kriging and conditional simulation of gaussian field. *Journal of engineering mechanics*, 121 (2):181–186.

Jones, D. R., Schonlau, M., and William, J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492.

Journel, A. and Kyriakidis, P. (2004). *Evaluation of Mineral Reserves: a Simulation Approach*. Oxford University Press, New York.

Matheron, G. (1973). The intrinsic random functions and their applications. *Advances in applied probability*, pages 439–468.

O'Hagan, A. (1978). Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40(1):1–42.

Omre, H. and Halvorsen, K. (1989). The bayesian bridge between simple and universal kriging. *Mathematical Geology*, 22 (7):767–786.

Roustant, O., Ginsbourger, D., and Deville, Y. (2012). Dicekriging, Diceoptim: Two R packages for the analysis of computer experiments by kriging-based metamodelling and optimization. *Journal of Statistical Software*, 51.

Santner, T. J., Williams, B. J., and Notz, W. (2003). *The Design and Analysis of Computer Experiments*. Springer Verlag.

Villemonteix, J., Vazquez, E., and Walter, E. (2009). An informational approach to the global optimization of expensive-to-evaluate functions. *Journal of Global Optimization*, 44(4):509–534.

Appendix A Universal and Simple Kriging

Let us consider a L^2 random field Z defined on a bounded set $\mathbb{X} \subset \mathbb{R}^d$ with known (and not necessarily stationary) covariance function $k(\cdot,\cdot)$ and unknown mean function $m(\cdot)$ such that $Z|m \sim \mathrm{GRF}(m,k)$, where $\mathrm{GRF}(m,k)$ denotes a Gaussian Random Field with mean function m and covariance function k. Let us apply a Bayesian approach and write m as follows:

$$m(\cdot) = \sum_{i=1}^{\ell} \beta_i f_i(\cdot)$$

where $\ell \geq 1$, $f_1,\ldots,f_\ell$ are ℓ known basis functions and $\beta = (\beta_1,\ldots,\beta_\ell)$ has an improper uniform prior in $\mathbb{R}^\ell$. In these settings, known as the Universal Kriging (UK) settings, when n observations $Z^{(n)} := Z(\mathbf{X}_n)$ are assimilated at points $\mathbf{X}_n := (\mathbf{x}_1,\ldots,\mathbf{x}_n) \in \mathbb{X}^n$, it is known (O'Hagan, 1978) that the posterior distribution of Z is a GRF with posterior (or conditional) mean function m_n^{UK} and covariance function k_n^{UK} given by the so-called UK equations:

$$\lambda^{UK}(\mathbf{x}) = K^{-1}\left(\mathbf{k}(\mathbf{x}) + \mathbb{F}(\mathbb{F}^\top K^{-1}\mathbb{F})^{-1}(\mathbf{f}(\mathbf{x}) - \mathbb{F}^\top K^{-1}\mathbf{k}(\mathbf{x}))\right) \tag{11}$$

$$m_n^{UK}(\mathbf{x}) = \lambda^{UK}(\mathbf{x})^\top Z^{(n)} = \mathbf{f}(\mathbf{x})^\top \widehat{\beta} + \mathbf{k}(\mathbf{x})^\top K^{-1}\left(Z^{(n)} - \mathbb{F}\widehat{\beta}\right) \tag{12}$$

$$\begin{aligned} k_n^{UK}(\mathbf{x},\mathbf{x}') = {} & k(\mathbf{x},\mathbf{x}') - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbf{k}(\mathbf{x}') + \\ & (\mathbf{f}(\mathbf{x})^\top - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbb{F})(\mathbb{F}^\top K^{-1}\mathbb{F})^{-1}(\mathbf{f}(\mathbf{x}')^\top - \mathbf{k}(\mathbf{x}')^\top K^{-1}\mathbb{F})^\top \end{aligned} \tag{13}$$

where $\widehat{\beta} := (\mathbb{F}^\top K^{-1}\mathbb{F})^{-1}\mathbb{F}^\top K^{-1} Z^{(n)}$, $\mathbf{f}(\mathbf{x}) := (f_1(\mathbf{x}),\ldots,f_\ell(\mathbf{x}))^\top$, $\mathbb{F} \in \mathbb{R}^{n\times\ell}$ is the matrix with row i equal to $\mathbf{f}(\mathbf{x}_i)^\top$, $\mathbf{k}(\mathbf{x}) := (k(\mathbf{x},\mathbf{x}_1),\ldots,k(\mathbf{x},\mathbf{x}_n))^\top$, K is the covariance matrix at the observation points, $K := (k(\mathbf{x}_i,\mathbf{x}_j))_{1\leq i,j\leq n}$. The vector $\lambda^{UK}(\mathbf{x})$ is the vector of n kriging weights of $\mathbf{x}_1,\ldots,\mathbf{x}_n$ for the prediction at point $\mathbf{x}$.

A well known simpler setting is the case where the non-conditional mean function m is already known. In that case, the Bayesian approach is no longer necessary and the conditional mean and covariance function of Z are given by the so-called Simple Kriging (SK) equations, written here in the case where $m(\cdot) = 0$:

$$\lambda^{SK}(\mathbf{x}) = K^{-1}\mathbf{k}(\mathbf{x}) \tag{14}$$

$$m_n^{SK}(\mathbf{x}) = \lambda^{SK}(\mathbf{x})^\top Z^{(n)} = \mathbf{k}(\mathbf{x})^\top K^{-1} Z^{(n)} \tag{15}$$

$$k_n^{SK}(\mathbf{x},\mathbf{x}') = k(\mathbf{x},\mathbf{x}') - \mathbf{k}(\mathbf{x})^\top K^{-1}\mathbf{k}(\mathbf{x}') \tag{16}$$

If m is not equal to zero, the SK covariance function k_n^{SK} is unchanged and an application of Eq. (15) to the centred GRF $Z - m$ yields $m_n^{SK}(\mathbf{x}) = m(\mathbf{x}) + \mathbf{k}(\mathbf{x})^\top K^{-1}(Z^{(n)} - m(\mathbf{x}^{(n)}))$.

Appendix B Algorithms

Algorithm 2 Preliminary cost (unfavorable case): computation of $Z^{(i)}(\mathbf{X}_q)$ for all $1 \leq i \leq M$.

Require: M i.i.d. GRFs $Z^{(1)},\ldots,Z^{(M)}$ simulated in p points $\mathbf{e}_1,\ldots,\mathbf{e}_p$ conditionally on $n \geq 0$ observations $Z(\mathbf{x}_1),\ldots,Z(\mathbf{x}_n)$ at points $\mathbf{x}_1,\ldots,\mathbf{x}_n$.
Require: The non-conditional covariance function, k, of the GRFs $Z^{(1)},\ldots,Z^{(M)}$.
Require: $q > 0$ additional points $\mathbf{X}_q = (\mathbf{x}_{n+1},\ldots,\mathbf{x}_{n+q})$.
Step 1. The goal is to compute a matrix of $(n+p)\times q$ kriging weights of $(\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{e}_1,\ldots,\mathbf{e}_p)$ for the prediction at points $\mathbf{x}_{n+1},\ldots,\mathbf{x}_{n+q}$, using Eq. (11).
- Compute the matrix K^{-1} of Eq. (11), i.e. the inverse covariance matrix at the $n+p$ points $(\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{e}_1,\ldots,\mathbf{e}_p)$: $\mathbf{O}((\mathbf{n}+\mathbf{p})^{\mathbf{3}})$
- Compute the other terms in Eq. (11) which do not depend on $\mathbf{x}$, i.e. $\mathbb{F}^\top K^{-1}$, then $(\mathbb{F}^\top K^{-1}\mathbb{F})^{-1}$, then $\mathbb{F}(\mathbb{F}^\top K^{-1}\mathbb{F})^{-1}$: $\mathbf{O}(\ell(\mathbf{n}+\mathbf{p})^{\mathbf{2}}+\ell^{\mathbf{2}}(\mathbf{n}+\mathbf{p})+\ell^{\mathbf{3}})$
for $i = 1 \to q$ **do**
- Compute $\mathbf{k}(\mathbf{x}_{n+i})$, i.e. the column vector $(k(\mathbf{x}_{n+i},\mathbf{x}_1),\ldots,k(\mathbf{x}_{n+i},\mathbf{x}_n),k(\mathbf{x}_{n+i},\mathbf{e}_1),\ldots,k(\mathbf{x}_{n+i},\mathbf{e}_p))^\top$ and $\mathbf{f}(\mathbf{x}_{n+i})$: $\mathbf{O}(\mathbf{n}+\mathbf{p}\ +\ \ell)$
- Compute the multiplication $K^{-1}\mathbf{k}(\mathbf{x}_{n+i})$ and then $\mathbb{F}^\top K^{-1}\mathbf{k}(\mathbf{x}_{n+i})$. Conclude the computation of $\lambda^{UK}(\mathbf{x}_{n+i})$: $\mathbf{O}(\ell(\mathbf{n}+\mathbf{p})+(\mathbf{n}+\mathbf{p})^{\mathbf{2}})$
end for
Total cost for Step 1: $\mathbf{O}((\mathbf{n}+\mathbf{p})^{\mathbf{3}}+(\mathbf{q}+\ell)(\mathbf{n}+\mathbf{p})^{\mathbf{2}}+(\ell^{\mathbf{2}}+\mathbf{q}\ell)(\mathbf{n}+\mathbf{p})+\ell^{\mathbf{3}})$
Total cost for Step 1 dominated by the term of $\mathbf{O}((\mathbf{n}+\mathbf{p})^{\mathbf{3}})$
Step 2. Preliminary: get the kriging covariance matrix $S := (k_{n+p}(\mathbf{x}_{n+i},\mathbf{x}_{n+j}))_{1\leq i,j\leq q}$ using Eq. (13) and the terms precomputed in Step 1. No new terms are added to the complexity (compared to Step 1), except a term of $\mathbf{O}(\mathbf{q}^{\mathbf{2}}(\mathbf{n}+\mathbf{p}))$ corresponding to q^2 vector-vector products. Compute also a decomposition (Cholesky, Mahalanobis) of S with a cost of $\mathbf{O}(\mathbf{q}^{\mathbf{3}})$.
for $i = 1 \to q$ **do**
- If $n > 0$, precompute, $u_i := \sum_{j=1}^{n}[\lambda^{UK}(\mathbf{x}_{n+i})]_j Z(\mathbf{x}_j)$, where $[\lambda^{UK}(\mathbf{x}_{n+i})]_j$ is the kriging weight of $\mathbf{x}_j$ for the prediction at point $\mathbf{x}_{n+i}$: $\mathbf{O}(\mathbf{n})$
end for
for $i = 1 \to M$ **do**
for $j = 1 \to q$ **do**
- Compute the kriging mean $m^{UK}_{n+p}(\mathbf{x}_{n+j})$ from the $n+p$ observations $Z(\mathbf{x}_1),\ldots,Z(\mathbf{x}_n),Z^{(i)}(\mathbf{e}_1),\ldots,Z^{(i)}(\mathbf{e}_p)$, using the vector of kriging weights $\lambda^{UK}(\mathbf{x}_{n+j})$. The precomputation of u_j reduces this to p operations. $\mathbf{O}(\mathbf{p})$.
end for
- Simulate a Gaussian random vector $Z^{(i)}(\mathbf{X}_q)$ with mean $m_{n+p}(\mathbf{X}_q)$ and covariance matrix S. $\mathbf{O}(\mathbf{q}^{\mathbf{2}})$
end for
Total cost for Step 2: $\mathbf{O}(\mathbf{M}(\mathbf{pq}+\mathbf{q}^{\mathbf{2}})+\mathbf{q}^{\mathbf{2}}(\mathbf{n}+\mathbf{p})+\mathbf{q}^{\mathbf{3}})$
Total cost for Step 2 dominated by the term of $\mathbf{O}(\mathbf{Mpq})$
Total cost of the algorithm dominated by the term of $\mathbf{O}(\mathbf{Mpq}+(\mathbf{n}+\mathbf{p})^{\mathbf{3}})$

Algorithm 3 Classical kriging residual algorithm

Require: M i.i.d. GRFs $Z^{(1)},\dots,Z^{(M)}$ simulated in $p+q$ points $\mathbf{e}_1,\dots,\mathbf{e}_p,\mathbf{x}_{n+1},\dots,\mathbf{x}_{n+q}$ conditionally on $n \geq 0$ observations $\mathbf{x}_1,\dots,\mathbf{x}_n$.
Require: The non-conditional covariance function, k, of the GRFs $Z^{(1)},\dots,Z^{(M)}$.
Require: $q > 0$ real observations $\mathbf{Z}_{\text{new}} = (Z(\mathbf{x}_{n+1}),\dots,Z(\mathbf{x}_{n+q}))$ at points $\mathbf{X}_q = (\mathbf{x}_{n+1},\dots,\mathbf{x}_{n+q})$.
Require: If $n > 0$, the inverse of the matrix $K_0 := (k(\mathbf{x}_i,\mathbf{x}_j))_{1\leq i,j\leq n}$

Step 1. The goal is to compute a matrix of $(n+q)\times p$ kriging weights of $(\mathbf{x}_1,\dots,\mathbf{x}_{n+q})$ for the prediction at points $\mathbf{e}_1,\dots,\mathbf{e}_p$, using Eq. (11). These weights will be used to obtain kriging means at points $\mathbf{e}_1,\dots,\mathbf{e}_p$.

- Compute the inverse of the $(n+q)\times(n+q)$ matrix K of Eq. (11), i.e. the covariance matrix K at the $n+q$ points $(\mathbf{x}_1,\dots,\mathbf{x}_{n+q})$. The inverse is obtained from the inverse of the $n\times n$ matrix K_0 using matrix block-inversion formulas $\mathbf{O(qn^2)}$.
- Compute the other terms in Eq. (11) which do not depend on $\mathbf{x}$, i.e. $\mathbb{F}^\top K^{-1}$, then $(\mathbb{F}^\top K^{-1}\mathbb{F})^{-1}$, then $\mathbb{F}(\mathbb{F}^\top K^{-1}\mathbb{F})^{-1}$: $\mathbf{O}(\ell(\mathbf{n+q})^\mathbf{2}+\ell^\mathbf{2}(\mathbf{n+q})+\ell^\mathbf{3})$

for $i = 1 \to p$ **do**
- Compute $\mathbf{k}(\mathbf{e}_i)$, i.e. the column vector $(k(\mathbf{e}_i,\mathbf{x}_1),\dots,k(\mathbf{e}_i,\mathbf{x}_{n+q}))^\top$ and $\mathbf{f}(\mathbf{e}_i)$: $\mathbf{O(n+q\ +\ \ell)}$
- Compute the multiplication $K^{-1}\mathbf{k}(\mathbf{e}_i)$ and then $\mathbb{F}^\top K^{-1}\mathbf{k}(\mathbf{e}_i)$. Conclude the computation of $\lambda^{UK}(\mathbf{e}_i)$: $\mathbf{O}(\ell(\mathbf{n+q})+(\mathbf{n+q})^\mathbf{2})$

end for

Total cost for Step 1: $\mathbf{O}((\mathbf{q}+\ell)\mathbf{n^2}+\ell\mathbf{q^2}+\ell^\mathbf{2}(\mathbf{n+q})+\ell^\mathbf{3}+\mathbf{p}((\mathbf{n+q})^\mathbf{2}+\ell(\mathbf{n+q})))$
Total cost for Step 1 dominated by the term of $\mathbf{O(p(n+q)^2)}$

Step 2. The goal is to use the computed kriging weights to perform, for all the M GRF sample paths, kriging means at points $\mathbf{e}_1,\dots,\mathbf{e}_p$ based on $n+q$ observations at points $\mathbf{x}_1,\dots,\mathbf{x}_{n+q}$.

for $i = 1 \to p$ **do**
- If $n > 0$, precompute, $u_i := \sum_{j=1}^n [\lambda^{UK}(\mathbf{e}_i)]_j Z(\mathbf{x}_j)$, where $[\lambda^{UK}(\mathbf{e}_i)]_j$ is the kriging weight of $\mathbf{x}_j$ for the prediction at point $\mathbf{e}_i$: $\mathbf{O(n)}$
- Compute the kriging mean at point e_j from the real $n+q$ observations $Z(\mathbf{x}_1),\dots,Z(\mathbf{x}_n),Z(\mathbf{x}_{n+1}),\dots,Z(\mathbf{x}_{n+q})$, using the vector of kriging weights $\lambda^{UK}(\mathbf{e}_i)$. The precomputation of u_j reduces this to q operations. $\mathbf{O(q)}$.

end for

for $i = 1 \to M$ **do**
  for $j = 1 \to p$ **do**
  - Compute the kriging mean at point e_j from the n real observations $Z(\mathbf{x}_1),\dots,Z(\mathbf{x}_n)$ and the q "artificial" observations $Z^{(i)}(\mathbf{X}_q) = (Z^{(i)}(\mathbf{x}_{n+1}),\dots,Z^{(i)}(\mathbf{x}_{n+q}))$. The precomputation of u_j reduces this to q operations. $\mathbf{O(q)}$.
  - Apply Eq. (1) to obtain $Z^{\star(i)}(\mathbf{e}_j)$

  end for

end for

Total cost for Step 2: $\mathbf{O(Mpq)}$
Total cost of the algorithm dominated by the term of $\mathbf{O(Mpq+p(n+q)^2)}$

Algorithm 4 FOXY algorithm based on Proposition 3

Require: M i.i.d. GRFs $Z^{(1)},\dots,Z^{(M)}$ simulated in $p+q$ points $\mathbf{e}_1,\dots,\mathbf{e}_p,\mathbf{x}_{n+1},\dots,\mathbf{x}_{n+q}$ conditionally on $n\geq 0$ observations $\mathbf{x}_1,\dots,\mathbf{x}_n$.
Require: The non-conditional covariance function, k, of the GRFs $Z^{(1)},\dots,Z^{(M)}$.
Require: $q>0$ real observations $\mathbf{Z}_{\text{new}}=(Z(\mathbf{x}_{n+1}),\dots,Z(\mathbf{x}_{n+q}))$ at points $\mathbf{X}_q=(\mathbf{x}_{n+1},\dots,\mathbf{x}_{n+q})$.
Require: If $n>0$, the inverse of the matrix $K_0:=(k(\mathbf{x}_i,\mathbf{x}_j))_{1\leq i,j\leq n}$
Step 1. The goal is to compute the q kriging weights $\lambda_{n,q}(\mathbf{x})=K_{n,q}^{-1}k_n^{UK}(\mathbf{x},\mathbf{X}_q)$ of Eq. (8) where $\mathbf{x}$ takes the values $\mathbf{e}_1,\dots,\mathbf{e}_p$, $K_{n,q}=(k_n^{UK}(\mathbf{x}_{n+i},\mathbf{x}_{n+j}))_{1\leq i,j\leq q}$ and k_n^{UK} is obtained using Eq. (13).
- Compute the inverse of the $n\times n$ matrix K of Eq. (16) i.e. the covariance matrix K at the n points $(\mathbf{x}_1,\dots,\mathbf{x}_n)$. This inverse is supposed to be precomputed here.
- Compute the other terms of Eq. (13) which do not depend on $\mathbf{x}$, i.e. $K^{-1}\mathbb{F}$, then $(\mathbb{F}^\top K^{-1}\mathbb{F})^{-1}$: $\mathbf{O}(\ell\mathbf{n^2}+\ell^\mathbf{2}\mathbf{n}+\ell^\mathbf{3})$
for $i=1\to p$ **do**
- Compute $\mathbf{k}(\mathbf{e}_i)$, i.e. the column vector $(k(\mathbf{e}_i,\mathbf{x}_1),\dots,k(\mathbf{e}_i,\mathbf{x}_n))^\top$ and $\mathbf{f}(\mathbf{e}_i)$: $\mathbf{O}(\mathbf{n}+\ell)$
- Compute the multiplication $\mathbf{k}(\mathbf{e}_i)^\top(K^{-1}\mathbb{F})$ and then $\mathbf{f}(\mathbf{e}_i)^\top-\mathbf{k}(\mathbf{e}_i)^\top(K^{-1}\mathbb{F})$: $\mathbf{O}(\ell\mathbf{n})$
end for
for $i=1\to q$ **do**
- Compute $\mathbf{k}(\mathbf{x}_{n+i})$, i.e. the column vector $(k(\mathbf{x}_{n+i},\mathbf{x}_1),\dots,k(\mathbf{x}_{n+i},\mathbf{x}_n))^\top$ and $\mathbf{f}(\mathbf{x}_{n+i})$: $\mathbf{O}(\mathbf{n}+\ell)$
- Compute the multiplication $\mathbf{k}(\mathbf{x}_{n+i})^\top(K^{-1}\mathbb{F})$ and then $\mathbf{f}(\mathbf{x}_{n+i})^\top-\mathbf{k}(\mathbf{x}_{n+i})^\top(K^{-1}\mathbb{F})$: $\mathbf{O}(\ell\mathbf{n})$
- Compute the multiplications $K^{-1}\mathbf{k}(\mathbf{x}_{n+i})$ and the multiplication $(\mathbb{F}^\top K^{-1}\mathbb{F})^{-1}(\mathbf{f}(\mathbf{x}_{n+i})^\top-\mathbf{k}(\mathbf{x}_{n+i})^\top K^{-1}\mathbb{F})^\top$: $\mathbf{O}(\mathbf{n^2}+\ell^\mathbf{2})$
end for
- Conclude the calculation of $k_n^{UK}(\mathbf{x},\mathbf{x}')$ for all $(\mathbf{x},\mathbf{x}')\in\mathbf{X}_q\times\mathbf{X}_q$: $\mathbf{O}(\mathbf{q^2}(\mathbf{n}+\ell))$
- Conclude the calculation of $k_n^{UK}(\mathbf{x},\mathbf{x}')$ for all $(\mathbf{x},\mathbf{x}')\in\{\mathbf{e}_1,\dots,\mathbf{e}_p\}\times\mathbf{X}_q$: $\mathbf{O}(\mathbf{pq}(\mathbf{n}+\ell))$
- Conclude the calculation the q kriging weights for prediction at points $\mathbf{e}_1,\dots,\mathbf{e}_p$ by computing the product: $k_n^{UK}(\mathbf{x},\mathbf{X}_q)^\top K_n^{-1}$ for all $\mathbf{x}\in\{\mathbf{e}_1,\dots,\mathbf{e}_p\}$: $\mathbf{O}(\mathbf{q^3}+\mathbf{pq^2})$
Total cost for Step 1: $\mathbf{O}(\mathbf{p}\ell\mathbf{n}+\mathbf{p}\ell\mathbf{q}+\mathbf{pqn}+\mathbf{pq^2}+\mathbf{n^2q}+\mathbf{q^2n}+\mathbf{q}\ell\mathbf{n}+\mathbf{q}\ell^\mathbf{2}+\ell\mathbf{n^2}+\ell^\mathbf{2}\mathbf{n}+\ell^\mathbf{3}+\mathbf{q^3})$
Total cost for Step 1 dominated by the term of $\mathbf{O}((\mathbf{pn}+\mathbf{n^2})(\ell+\mathbf{q}))$
Step 2. Conclude the update of the M GRFs by applying Eq. (8). For a given point $\mathbf{x}$ and a given GRF sample path, q operations are performed. Total cost of $\mathbf{O}(\mathbf{Mpq})$.
Total cost of the algorithm dominated by the term of $\mathbf{O}(\mathbf{Mpq}+(\mathbf{pn}+\mathbf{n^2})(\ell+\mathbf{q}))$