



HAL
open science

An Automatic Workflow for Risk Analysis on Spatial output Properties using Kriging-based Surrogate Models -Application to Stratigraphic forward Modeling

Véronique Gervais, Didier Granjeon, Sarah Bouquet

► To cite this version:

Véronique Gervais, Didier Granjeon, Sarah Bouquet. An Automatic Workflow for Risk Analysis on Spatial output Properties using Kriging-based Surrogate Models -Application to Stratigraphic forward Modeling. Basin Research, 2023, 35 (5), pp.1933-1960. 10.1111/bre.12786 . hal-04582862

HAL Id: hal-04582862

<https://ifp.hal.science/hal-04582862v1>

Submitted on 22 May 2024

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

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

An automatic workflow for risk analysis on spatial output properties using kriging-based surrogate models - Application to stratigraphic forward modeling

Véronique Gervais*, Didier Granjeon, Sarah Bouquet

IFP Energies Nouvelles
1 & 4 avenue de Bois-Préau
92852 Rueil-Malmaison Cedex
France

*corresponding author – veronique.gervais@ifpen.fr

Abstract

Stratigraphic numerical forward process-based models represent the formation and evolution of sedimentary basins through time. Their main deliverable is a 3D digital grid which can help to better understand the sedimentary basin infill. These models depend on several input parameters that need to be characterized for the studied basin. However, available data such as well logs and seismic data may not provide enough information to identify a unique possible value for these parameters. It is then crucial to take the uncertainty induced by this non-uniqueness into account in the decision-making process. As a single numerical stratigraphic forward simulation can be very time consuming, solutions are needed to limit the computational cost required to estimate such uncertainties. In particular, machine learning techniques can be used to build meta-models that mimic the simulator and provide fast estimations of its outputs for any values of the input parameters. The key step for the efficiency of such an approach stands in the choice of the set of simulations (or training set) used to build the meta-models: it should be informative enough to obtain accurate predictions for the output properties of interest, but also of reasonable size to limit simulation time. Then, meta-models can be used to investigate a large number of models and make uncertainty quantification easier. Here, we focus on the prediction of spatial outputs of interest approximated from the joint use of several kriging-based meta-models combined to reduced-basis decomposition. Sequential approaches have been proposed in the literature to identify training sets iteratively for a kriging-based meta-model, building upon the specific structure of such surrogates. We propose here to extend these approaches to the context of spatial output predictors. The results obtained on two synthetic test cases, representing a carbonate platform and a clastic environment, highlight the potential of the proposed approach for risk analysis to iteratively build training sets with a satisfactory efficiency in terms of simulation time and prediction accuracy.

Highlights

- Use of machine learning techniques for risk analysis on numerical model spatial outputs
- Iterative definition of the training set used to build kriging-based predictors for spatial output properties
- Assessment of the method potential for stratigraphic forward modeling on two synthetic cases
- Validation cases representing a carbonate platform and a clastic environment
- Use of the resulting surrogate models in uncertainty studies for the test cases spatial outputs

1. Introduction

Stratigraphic numerical forward process-based models represent the formation and evolution of sedimentary basins through time [25]. Their main deliverable is a 3D digital grid which can help to better understand the sedimentary basin infill and characterize for example the spatial variation of sedimentary layer thickness and properties. However, these models depend on several input parameters that need to be characterized for the studied basin (*e.g.*, subsidence map, sea or lake level curve, transport process parameters, sediment source locations ...). Available data such as well logs and seismic can be used to help in this characterization. However, they may not provide enough information to identify a unique combination of values for the parameters, resulting in several acceptable models that may fit the data, but correspond to different dynamic evolutions and provide different solutions away from the wells [33]. It is then crucial to quantify the residual uncertainty induced by this non-uniqueness and to take it into account in the decision-making process, for instance through the estimation of the distribution in the basin of the probability of occurrence of sediment facies controlling reservoir presence. However, due to the potentially large number of uncertain parameters and the high computational cost of performing numerical stratigraphic forward simulations (a single high-resolution simulation can require a few hours to run on a cluster), it can be difficult to reduce and estimate thoroughly uncertainties in operational studies.

A way to reduce the computational cost of the uncertainty quantification process consists in using machine learning techniques to build meta-models (also called surrogates, emulators, or proxy models) that mimic the simulator and provide very fast estimations of its outputs for any values of the input parameters. More precisely, a meta-model is a parameterized function that approximates the relationship between the simulator input parameters and a given output property. Different techniques can be used to build such meta-models, which differ from the structure of the function used. They include for instance polynomial regression, Gaussian processes (rooted in kriging) and neural networks (see also [9] for a review). Once chosen the set of input parameters to be considered, their target ranges of variation and the structure of the surrogate, the meta-model parameter values are identified through the minimization of a cost function that quantifies the mismatch between a set of simulated values for the output to be approximated (the training set) and the corresponding predictions. These parameters are for instance the weights associated to each term in polynomial regression. For Gaussian process-based meta-modeling, the predictions are equal to the kriging estimate of the output on the training set, and the meta-model parameters include the mean and covariance function characteristics [28]. Then, if the predictions provided by the meta-model for any input parameter values are close to the corresponding simulated values, they can be used instead of the simulator. Meta-modeling thus appears as an attractive tool for uncertainty management as long as the number of simulations required to obtain accurate predictions is reasonable. Indeed, it makes it possible to investigate rapidly a huge number of models, and to perform statistical analyses such as quantitative sensitivity analysis or uncertainty propagation. It can thus help to better understand and quantify the impact of the input uncertainty on the dynamic processes. Meta-models can also be combined to minimization algorithms to identify models reproducing available data. Note however that meta-models are built for a given output and set of input parameters and are only valid for the uncertainty ranges chosen for these parameters.

The choice of the training set appears as a key step for the efficiency of the meta-modeling approach, especially for very long simulations. Indeed, this ensemble should be small to limit computation times, but also sufficiently informative to obtain accurate estimations of the output of interest for any input parameter values within the range of uncertainty. In practice, it appears difficult to generate in one trial a training set that is optimal in terms of both simulation time and prediction accuracy. Sequential techniques have thus been proposed in the literature to build training sets iteratively. They consist in generating first a small sample of the parameter space, and in complementing this ensemble sequentially. The idea behind is to stop the process as soon as a satisfactory prediction accuracy has been reached, and thus avoid unnecessary simulations. Various criteria have been

investigated to choose the next point of the parameter space to be added to the training set in order to improve as fast as possible the meta-model accuracy. Some of them act in the parameter space only and are based on geometrical considerations. For instance, a maximin criterion is considered in [19]. Other techniques are dedicated to kriging-based meta-models and build upon their specific structure. In particular, several of these approaches use the kriging variance that represents the model mean square error (MSE) and provides some quantification of the uncertainty on the predictions. These techniques then aim at identifying points that should reduce this uncertainty as fast as possible in the whole parameter space, and thus improve rapidly the prediction accuracy. For instance, the new point added to the training set can be the one that maximizes the kriging variance [28]. Variants narrow the search space to specific regions of the parameter space [3,31] or weight the kriging variance by an indicator on the meta-model accuracy [20]. Another interesting criterion is the integrated MSE (IMSE) obtained by integration of the kriging variance over the parameter space [28]: the next point to be simulated is chosen as the one that should reduce the most the IMSE given the current meta-model characteristics. These different criteria are referred to as adaptive as they depend on the characteristics of the current meta-model. Note that other approaches aim at improving the prediction accuracy in some specific regions of the parameter space, *e.g.*, corresponding to low values of the error on the available data for calibration [18]. However, they are beyond the scope of this paper.

The adaptive approaches described previously can be directly applied to kriging-based meta-models approximating scalar outputs such as the mean sediment thickness or the average sand concentration. However, for many fields of application, and in particular stratigraphic and basin modeling, the value of a given output property in all grid blocks or columns of the grid can play a key role in the decision-making process. For such spatially dependent properties, a Proper Orthogonal Decomposition (POD) [21] can be considered first to limit the number of variables to be approximated, and thus the computation times. More precisely, this technique makes it possible to identify a limited number of latent output variables (much less than the number of grid blocks from our experience) that reproduce the training set when appropriately weighted. The meta-modeling then applies to these new variables, *e.g.*, using kriging. More details can be found in [22,10] for instance. This workflow was applied to perform sensitivity and risk analysis on various spatial output properties, both in stratigraphic forward modeling (clastic environment [10], organic matter production and preservation [2]) and in basin modeling (source rock potential [8], biogenic gas generation [32]). However, in these studies, the definition of a training set that is optimal in terms of both computation time and accuracy is not clearly addressed. The objective of the current work is thus to simplify the choice of the training set by proposing an automatic sequential design strategy for the context of spatial outputs, in which several meta-models are considered simultaneously to predict the output property on the whole set of grid blocks at once. To achieve this, we focus on the adaptive criteria described above for a given kriging-based meta-model and we propose a way to combine them with the risk analysis workflow for spatial output properties. The efficiency of the resulting approach for iteratively building training sets that provide accurate predictions of such spatial properties in the whole parameter space will be assessed in the context of stratigraphic forward modeling on two synthetic cases, representative of a carbonate platform and a clastic environment. Note that the proposed workflow applies similarly to predict properties that depend on time or depth. Then, the use of the resulting predictors will be illustrated for sensitivity and risk analysis. The calibration process will be slightly addressed through preliminary studies. However, the use of meta-modeling in the context of assisted calibration is beyond the scope of this study dedicated to spatial outputs and requires specific design strategies (*e.g.*, EGO [18]).

The paper outlines as follows. The sequential kriging-based workflow proposed in this work is described in section 2, while section 3 presents the results obtained in stratigraphic forward modeling, considering a clastic and a carbonate environment. Different criteria for the choice of the next points added to the training set are compared, and the resulting predictors are used to perform various uncertainty studies to highlight the potential of the approach for geologists and sedimentologists.

2. Meta-modeling workflow

The proposed sequential workflow for risk analysis on spatial outputs is schematized in Figure 1 and summarized below. It only differs from the one used in [10,2,8,32] by the additional step 5 corresponding to the process of complementing the initial sample. More details on the successive steps are given in the following subsections and in the Appendix.

Algorithm 1

1. Identify the model input parameters that need to be considered in the uncertainty study (these parameters are assumed independent); assign them a range of plausible values and a probability distribution (uniform by default)
2. Generate an initial sample of the parameter space and perform direct simulations for the models of this ensemble
3. Extract the spatial output of interest from these simulations to form the training set
4. Build a predictor for the spatial output of interest using the training set and an approach combining Proper Orthogonal Decomposition (POD) and kriging-based meta-modeling
5. While the quality of the estimator for the property of interest is not satisfactory,
 - a. Complement the training set with new simulations chosen according to a pre-defined criterion applied to the current predictor
 - b. Build a predictor for the output of interest using the augmented training set
6. Use the final predictor to investigate new models, for instance to perform sensitivity and risk analysis

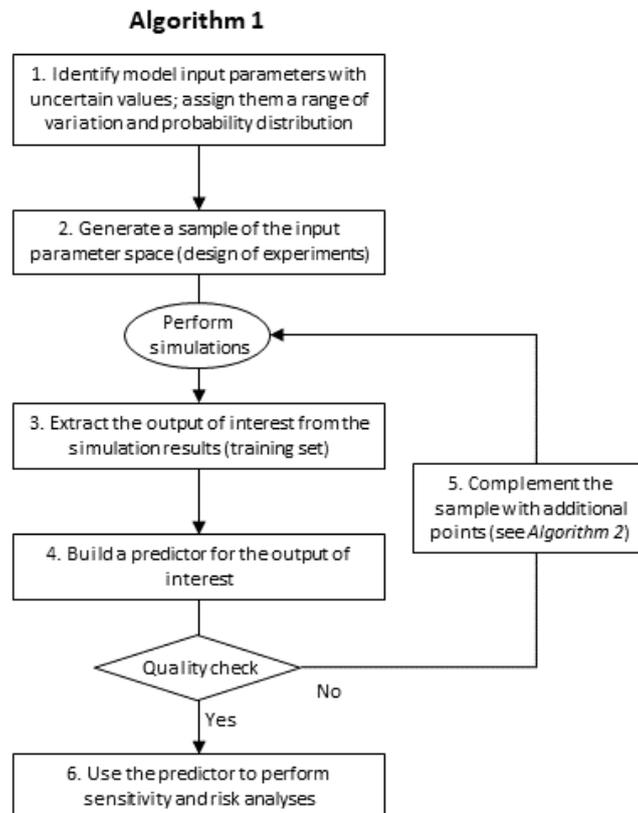


Figure 1 – General sequential workflow

2.1 Initial sampling

We assume identified a set of continuous input parameters $\theta \in \mathbb{R}^d$ to be considered in the uncertainty study, with ranges of variation defining the subspace $\Omega \subset \mathbb{R}^d$ in which the study needs to be performed (step 1). The choice of these intervals can be based for instance on the geologist knowledge, on bibliographical studies or on laboratory experiments. Then a probability law different from the uniform one can be assigned to each parameter if some values within the chosen range are considered more likely than others (e.g., normal or lognormal distributions). In addition, a log-transformation can be applied to parameters whose range of variation extends over several orders of magnitude to ensure a good coverage of the lower values.

An initial sample of Ω (the design of experiments) is then required to start the process (step 2). Here, we use the Latin Hypercube Sampling approach [28] to generate this initial design of experiments, denoted $\mathbf{D} = \{\theta_1, \dots, \theta_n\}$, $\theta_i \in \mathbb{R}^d$. With this technique, the size n of the generated sample is chosen first. Then, the range of variation of each parameter is divided into n equiprobable intervals with regard to its probability law. Finally, the n points retained in the design are generated randomly in Ω and satisfy the following property: only one point is located within each equiprobable interval for each parameter (see Figure 2 for illustration). The LHS approach provides a good coverage of the uncertainty space for a given size n , with points homogeneously distributed in the variation intervals for each parameter with regard to their probability law. The a priori knowledge is thus taken into account, leading to more dense samples in more probable areas. Note however that other sampling techniques could be considered as well to generate the initial sample.

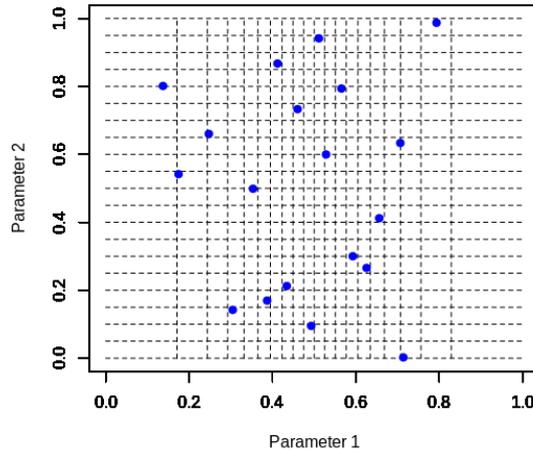


Figure 2 - Example of LHS of size 20 for two parameters. Parameter 1 follows a normal law with mean 0.5 and variance 0.04. Parameter 2 follows a uniform law and both parameters vary between 0 and 1. The blue dots indicate the sampling points. The dotted black lines separate the 20 equiprobable intervals for the two parameters.

2.2 Meta-modeling approach for spatial outputs

Once the simulations have been performed for the models of sample \mathbf{D} , the relationship between the input parameters θ and the output properties of interest can be approximated with meta-models. Let us denote $y(\theta, \mathbf{x})$ a target spatial output of interest, where \mathbf{x} represents the spatial location. In stratigraphic forward modeling, y can refer for instance to the distribution of the deposited sediment thickness or lithology concentration in each column of the grid. We assume known the values of y for sample \mathbf{D} and a set of locations in the spatial domain $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$. These values, denoted by $\mathbf{y}_D \in \mathbb{R}^{N \times n}$ with $(\mathbf{y}_D)_{ij} = y(\theta_j, \mathbf{x}_i)$, form the training set (step 3 of Algorithm 1). Then, following the approach described in [22,10], for any point of the parameter space $\theta^* \in \mathbb{R}^d$, we consider the estimator of y at location \mathbf{x}_i , $i \in \{1 \dots N\}$, defined using \mathbf{y}_D by:

$$\hat{y}(\theta^*, \mathbf{x}_i) = \bar{y}(\mathbf{x}_i) + \sum_{k=1}^L \hat{\alpha}_k(\theta^*) \phi_k(\mathbf{x}_i) \quad (1)$$

$\bar{y}(\mathbf{x}_i)$ denotes the average value of the observations at \mathbf{x}_i : $\bar{y}(\mathbf{x}_i) = \frac{1}{n} \sum_{j=1..n} y(\boldsymbol{\theta}_j, \mathbf{x}_i)$. The basis vectors $\phi_k(\mathbf{X}) \in \mathbb{R}^N$ considered in this linear formulation result from the use of Proper Orthogonal Decomposition (POD) on the data set \mathbf{y}_D [21]. $\hat{\alpha}_k$ denotes a meta-model approximating component α_k , the projection coefficient of y on ϕ_k . It is built using the set of projection coefficients on ϕ_k for the training set \mathbf{y}_D , denoted $\boldsymbol{\alpha}_k^D$. The basis vectors are sorted by decreasing order of explained variance, and only the L first are retained in the formulation to limit computation times. They correspond to a given percentage of cumulated variance (e.g., 95%), chosen as a trade-off between the number of meta-models to be computed and the loss of information.

Each component α_k , $k = 1, \dots, L$ is assumed to be a realization of a Gaussian process with a given mean and covariance function. It is estimated by the expectation of this Gaussian process conditioned to the data points $\boldsymbol{\alpha}_k^D$, corresponding to the kriging estimate of α_k on $\boldsymbol{\alpha}_k^D$. The kriging variance \hat{s}_k^2 characterizes the variance of the conditional Gaussian process and represents the model mean square error. It equals zero at the sampling points \mathbf{D} . More details on this approach can be found in [28] and in the Appendix. Finally, the definition of the predictor $\hat{\alpha}_k$ requires estimating the Gaussian process covariance function and mean. This is usually done by maximizing the logarithm of the likelihood function as described for instance in [29]. Here, we use the R package DiceKriging [27] to compute the kriging-based meta-models.

Let us note that, for a given covariance function, the kriging variance does not depend on the actual value of the target output for the sample, but only on the location of the sampling points in the parameter space. It is thus straightforward to compute its value for any sample \mathbf{D} without the need to estimate $\boldsymbol{\alpha}_k^D$. This property is a building block in the identification of a batch of points in sequential design strategies (see also [20] for instance).

The meta-models defined by (1) do not ensure that the predictions will range within physical values, in particular nonnegative ones for thickness or concentrations. In what follows, we thus apply the kriging approach (1) to a transformed function for some properties, namely their square root $z = \sqrt{y}$. The predictions for y are then taken equal to $\hat{y} = (\hat{z})^2$, what ensures nonnegative values.

2.3 Quality check

The predictor \hat{y} defined in equation (1) needs to be sufficiently accurate to be used instead of the simulator in the risk analysis process. In other words, its estimations must be close to the corresponding true simulated values for any $\boldsymbol{\theta}^* \in \Omega$. Different indices are generally used to check this accuracy, referring to the computation of the deviation between a set of simulated values and the corresponding estimations provided by \hat{y} .

If the user can afford it, it is recommended to generate an additional sample of the parameter space $\mathbf{D}_{test} = \{\tilde{\boldsymbol{\theta}}^1, \dots, \tilde{\boldsymbol{\theta}}^{n_{test}}\}$, $\tilde{\boldsymbol{\theta}}^i \in \Omega$, independent of the training set \mathbf{D} , and to perform the corresponding simulations. Then, the overall predictivity of \hat{y} can be assessed at each location \mathbf{x}_i , $i \in \{1 \dots N\}$ with the $R2$ coefficient of determination as:

$$R2(\hat{y}(\mathbf{x}_i)) = 1 - \frac{\sum_{j=1}^{n_{test}} (y(\tilde{\boldsymbol{\theta}}^j, \mathbf{x}_i) - \hat{y}(\tilde{\boldsymbol{\theta}}^j, \mathbf{x}_i))^2}{\sum_{j=1}^{n_{test}} (y(\tilde{\boldsymbol{\theta}}^j, \mathbf{x}_i) - \bar{y}(\mathbf{x}_i))^2} \quad (2)$$

where $\bar{y}(\mathbf{x}_i)$ represents the mean of $(y(\tilde{\boldsymbol{\theta}}^j, \mathbf{x}_i))_{1 \leq j \leq n_{test}}$. The $R2$ index refers to the sum of the squared errors on the predictions normalized by the sum of the squared deviation from the mean. It gets closer to 1 when prediction errors decrease, thus indicating more accurate estimations. If it is close to zero, the predictor is not more accurate on average than a meta-model that always predicts the mean $\bar{\alpha}_k$. This index can be considered in particular as a stopping criterion for the sequential process, using for instance a target minimal average value over \mathbf{X} : $\frac{1}{N} \sum_{i=1..N} R2(\hat{y}(\mathbf{x}_i)) > \beta$. In addition, the median value $R2_m$ of the $R2$ coefficients in all grid blocks will be considered in section 3 to make overall comparisons between the various experiments:

$$R2_m(\hat{y}) = \text{median} (R2(\hat{y}(\mathbf{x}_i)))_{i=1..N} \quad (3)$$

Note that the $R2$ index can also be computed for the component estimators $\hat{\alpha}_k, k = 1..L$ (see Appendix). In what follows, the test set \mathbf{D}_{test} will be generated using the LHS approach. The distribution of the quality check sample will thus be in accordance with the one used afterwards in the risk analysis process.

Other quality criteria can be computed from the training set itself, avoiding additional simulation cost. We consider such an index for the component predictors in the sequential process. More specifically, we refer to the leave-one-out cross-validation technique (LOO-CV) that consists in computing additional meta-models using $n - 1$ points of the training set, and in considering the prediction error on the left-out point to assess the meta-model predictivity. Denoting $\hat{\alpha}_{k,-j}(\boldsymbol{\theta})$ the kriging estimator computed from the training set $(\alpha_k(\boldsymbol{\theta}^i))_{1 \leq i \leq n, i \neq j}$, the $Q2$ coefficient builds upon the cross-validation error $\alpha_k(\boldsymbol{\theta}^j) - \hat{\alpha}_{k,-j}(\boldsymbol{\theta}^j)$ at point $\boldsymbol{\theta}^j$ as

$$Q2(\hat{\alpha}_k) = 1 - \frac{\sum_{j=1}^n (\alpha_k(\boldsymbol{\theta}^j) - \hat{\alpha}_{k,-j}(\boldsymbol{\theta}^j))^2}{\sum_{j=1}^n (\alpha_k(\boldsymbol{\theta}^j) - \bar{\alpha}_k)^2} \quad (4)$$

where $\bar{\alpha}_k$ represents the mean of $(\alpha_k(\boldsymbol{\theta}^j))_{1 \leq j \leq n}$. The $Q2$ coefficient gets closer to 1 when cross-validation errors decrease. In practice, some characteristics of $\hat{\alpha}_k$ such as the covariance function identified from the complete training set can be used to build meta-models $\hat{\alpha}_{k,-j}$ and limit the additional computational cost. More details can be found for instance in [7]. Note that in the following, the normalized cross-validation squared error $e_{cv}(\boldsymbol{\theta}^j)$ for point $\boldsymbol{\theta}^j \in \mathbf{D}$ refers to:

$$e_{cv}(\boldsymbol{\theta}^j) = \frac{(\alpha_k(\boldsymbol{\theta}^j) - \hat{\alpha}_{k,-j}(\boldsymbol{\theta}^j))^2}{\hat{s}_{k,-j}^2(\boldsymbol{\theta}^j)} \quad (5)$$

with $\hat{s}_{k,-j}^2$ the kriging variance associated to meta-model $\hat{\alpha}_{k,-j}$.

2.4 Choice of additional points for the design of experiments

2.4.1 General approach

If the predictor \hat{y} is not accurate enough, in particular with respect to the quality indices introduced above, a way to improve its accuracy consists in complementing the training set with additional simulations. As detailed in the introduction, several approaches have already been proposed in the literature to perform such active learning for a kriging-based meta-model (see [20] for instance for a review). The specificity of our problem lies in the simultaneous use of several meta-models $\hat{\alpha}_k$ to predict property y at all locations \mathbf{X} at once. To combine existing adaptive approaches to this configuration, we propose to apply them sequentially to components α_k from 1 to L , *e.g.*, in decreasing order of explained variance. More precisely, at each iteration, we identify the components which are not approximated by the current meta-models with sufficient accuracy, and we apply an adaptive criterion to the one among them having the smallest index. The idea behind is to reach first a good approximation for the main features of the spatial outputs, before considering more detailed ones. The resulting approach is described below (Algorithm 2). It corresponds to step 5 of Algorithm 1. The proposed workflow also includes the possibility to identify q points at a time at each iteration (*i.e.*, without running intermediary simulations), what appears as a key functionality if several simulations can be performed simultaneously, for instance on a computer cluster. The definition of the adaptive criterion used to identify the next points to be simulated is discussed in the next section (2.4.2).

Algorithm 2

- 1) Run through components α_k in decreasing order of explained variance, or equivalently from 1 to L , and identify the first one k^* for which the corresponding meta-model $\hat{\alpha}_{k^*}$ is not accurate enough.
- 2) Based on meta-model $\hat{\alpha}_{k^*}$ characteristics, identify q new points $\boldsymbol{\theta}_{new}^1, \dots, \boldsymbol{\theta}_{new}^q$ to be simulated, and add them to the design $\mathbf{D} \leftarrow \mathbf{D} \cup (\boldsymbol{\theta}_{new}^1, \dots, \boldsymbol{\theta}_{new}^q)$
- 3) Perform the simulations for $\boldsymbol{\theta}_{new}^1, \dots, \boldsymbol{\theta}_{new}^q$ and complement the training set: $\mathbf{y}^D \leftarrow \mathbf{y}^D \cup \{y(\boldsymbol{\theta}_{new}^1, \mathbf{X}), \dots, y(\boldsymbol{\theta}_{new}^q, \mathbf{X})\}$

- 4) Compute the POD of the updated training set \mathbf{y}^D and build kriging-based estimators for the resulting components α_k using \mathbf{y}^D . Finally, form the new estimator $\hat{\mathbf{y}}$ for the target output using (1).

If $q > 1$, a variant of this algorithm consists in rather identifying one point for each of the q components with the smallest indices that do not have accurate enough predictors. Also note that, as the POD is recomputed for each updated training set, components α_k and basis vectors ϕ_k may differ from one iteration to the next. In addition, this approach boils down to the classical adaptive approach described in previous works for a single discretization point ($N = 1$).

The quality criterion considered in Algorithm 2 to identify the meta-model $\hat{\alpha}_{k^*}$ to be improved can be defined as a target value for the meta-model accuracy, represented by the cross-validation $Q2$ coefficient (4) or the $R2$ coefficient. In what follows, we consider a target $Q2$ value of 0.95.

Finally, if several output properties need to be approximated in the study, Algorithm 2 can be considered to complement the training set for each of these properties alternatively from one iteration to the next, or to identify points related to each property at each iteration. In what follows, we rather focus on a single output property in the sequential process, and then study the potential of the resulting design of experiments to accurately predict other properties, or to be used as a starting point to improve predictions on these properties if necessary.

2.4.2 Strategies for the identification of new points

As mentioned in the introduction, several criteria have been proposed in the literature to complement a training set for a kriging-based meta-model. In this work, we focus on three of them to assess the potential of the proposed workflow. They are described in the next sections, where $\hat{\alpha}_k$ denotes the kriging-based estimator to be improved and \hat{s}_k^2 its variance.

Maximum Mean Square Error (MMSE)

The first criterion considered in this study deals with the Mean Square Error (MSE) corresponding to the kriging variance. We consider here that this variance $\hat{s}_k^2(\boldsymbol{\theta})$ at $\boldsymbol{\theta} \in \Omega$ provides some quantification of the uncertainty on the prediction $\hat{\alpha}_k(\boldsymbol{\theta})$. The next point $\bar{\boldsymbol{\theta}}$ added to the training set is then the one for which the uncertainty approximated by the kriging variance is the largest in the parameter space: $\hat{s}_k^2(\bar{\boldsymbol{\theta}}) = \max_{\boldsymbol{\theta} \in \Omega} \hat{s}_k^2(\boldsymbol{\theta})$. Here, $\bar{\boldsymbol{\theta}}$ is searched among a Monte Carlo sample of the parameter space of size 10000. However, Sobol' sequences or optimization approaches could also be envisioned in future work (see [20] for instance).

For a given covariance function, the kriging variance \hat{s}_k^2 only depends on the location of the sampling points in the parameter space (see Appendix). This makes it possible to identify several new points in a single iteration as described in [20] before running the corresponding simulations.

This strategy is illustrated in Figure 3. The reference 1D function to be approximated is plotted in blue. The training set initially consists of the 4 points highlighted by blue circles. The corresponding initial kriging-based predictor is given by the dotted blue line and its associated variance by the black curve. This variance exhibits null values at the sampling points and increases between them. The point $\bar{\boldsymbol{\theta}}$ with the largest kriging variance (or MSE) is indicated by the red star in the left figure. Complementing the training set with this point leads to the updated kriging variance plotted in red in the middle graph. This variance now also exhibits a null value at $\bar{\boldsymbol{\theta}}$ and equals the initial MSE for points distant enough from $\bar{\boldsymbol{\theta}}$. The new point corresponding to the largest value of the updated variance is again highlighted by the red star. The kriging variance resulting from the integration of this second point in the sample is plotted in red in the third graph. Note that here, the predictor (blue dotted line) is not updated as the objective is to illustrate the identification of several points in the same iteration.

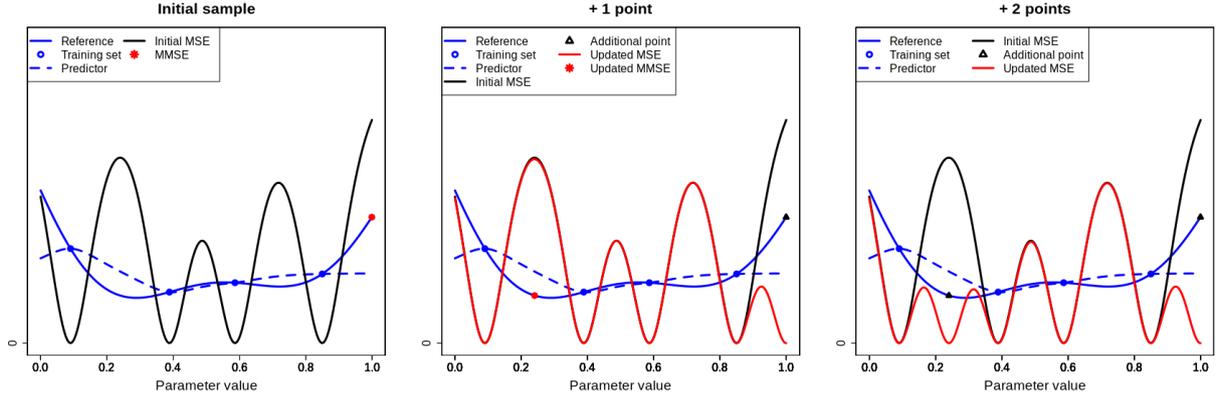


Figure 3 – Adaptive design strategy based on the MMSE criterion for a 1D function: update of the kriging variance when adding two points to the sample during the same iteration

Local MMSE

With the previous approach, the next points are searched for in the whole parameter space Ω . Authors in [3,31] rather propose to narrow this search space. Here, we refer to the approach introduced in [31]. First, the parameter space Ω is partitioned into n regions corresponding to the Voronoi polygons of the sample \mathcal{D} . Then, the Voronoi cell Ω^I associated to the point θ^I of the training set for which the normalized cross-validation squared error (5) is the largest is identified: $e_{CV}(\theta^I) = \max_{\theta^i \in \mathcal{D}} e_{CV}(\theta^i)$. Finally, the next point $\bar{\theta}$ added to the sample is chosen in Ω^I according to the MMSE criterion, or equivalently as the point for which the kriging variance is the largest in Ω^I . The objective is to better integrate the predictor accuracy in the sequential process. This approach is referred to as the Local Maximum Mean Square Error (L-MMSE) in what follows. As for the global MMSE criterion, $\bar{\theta}$ is searched for among a large sample of Ω^I that is generated in two steps following [31]: first, a Sobol' sequence is generated in the parameter space Ω ; then, the points of the sample that are within the subregion Ω^I are identified and stored. This is repeated until a large enough number of points has been found in Ω^I . However, in practice, this process can become very long after some iterations, and we resorted for a few of the experiences presented in section 3 to loosen the constraint on the Ω^I sampling size to keep reasonable computation times.

To select q points at a given iteration, the MMSE criterion is used to identify one point in each of the q first Voronoi cells sorted by decreasing order of the associated normalized cross-validation squared error e_{CV} . The kriging variance is updated as previously each time a new point is selected. This is illustrated in Figure 4 using the same 1D function as above. The Voronoi cell associated to the point with the largest cross-validation error is highlighted in grey in the left graph. The first point added to the sample is then the one with the largest kriging variance in this subspace (red star) and differs here from the one with the largest MMSE value in Ω . The resulting updated kriging variance is plotted in red in the middle figure, where the grey region corresponds to the Voronoi cell associated to the point of the training set with the second largest error e_{CV} .

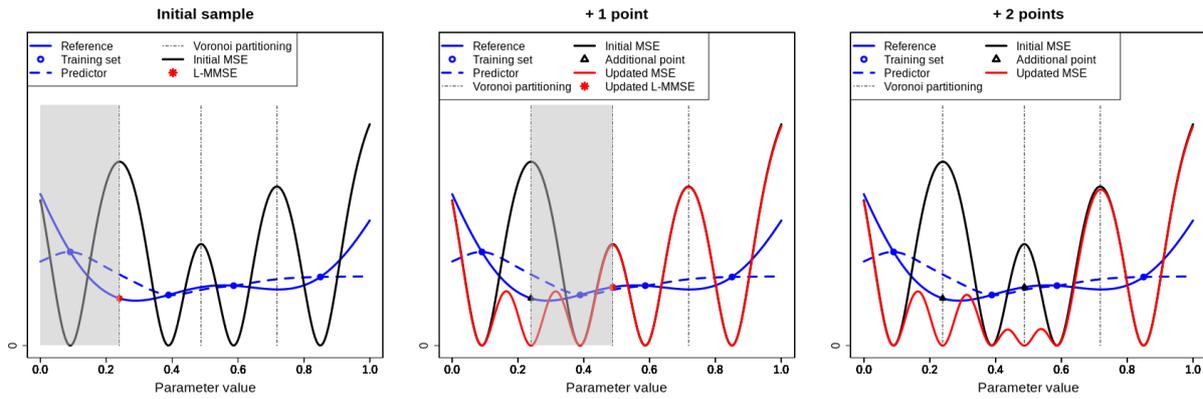


Figure 4 – Adaptive design strategy based on the L-MMSE criterion for a 1D function: update of the kriging variance when adding two points to the sample during the same iteration

Integrated Mean Square Error (IMSE)

The last criterion considered in this paper refers to the integrated value of the MSE over the parameter space, referred to as IMSE. The IMSE corresponds more precisely to the area of the region located below the MSE curve and highlighted in grey in Figure 5 for the same 1D function as above. The adaptive IMSE criterion then consists in identifying the point of the parameter space that, if added to the sample, would induce the largest IMSE reduction (or equivalently the smallest updated IMSE). For each candidate point $\theta^* \in \Omega$, the kriging variance corresponding to the sample $\mathbf{D} \cup \theta^*$ is computed with the current covariance function, enabling the estimation of the corresponding IMSE. This is illustrated in Figure 5 for two candidate points denoted by red stars (graphs in the middle and on the right). For each of these points, the corresponding updated MSE is given in red, highlighting the resulting reduction of the IMSE. As for the two previous approaches, the process can be repeated to identify sequentially several points to be added to the training set before running the corresponding simulations.

In what follows, the point that minimizes the IMSE is identified using the KrigInv R package [26]. The IMSE is estimated using a Sobol' sequence of size $100 \times d$, and the minimization problem is solved using the Genoud evolutionary optimization algorithm (GENetic Optimization Using Derivatives).

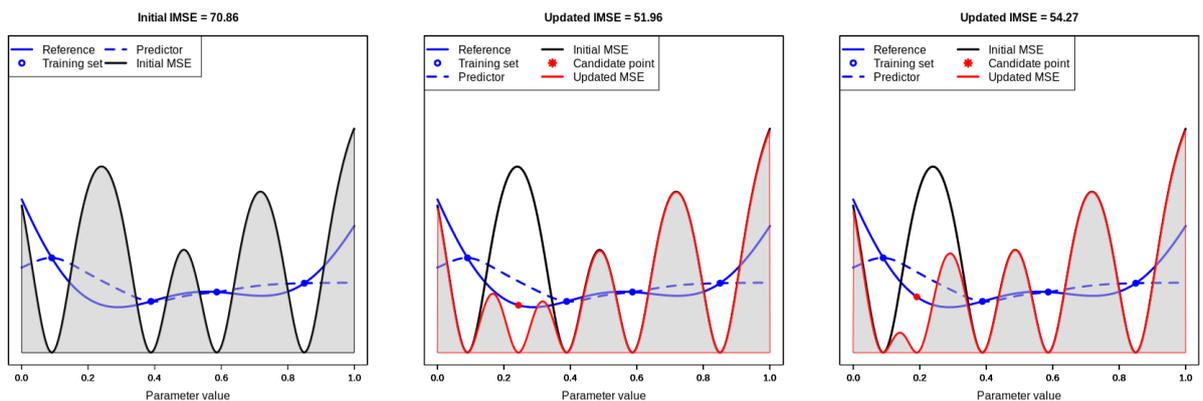


Figure 5 – Adaptive design strategy based on the IMSE criterion for a 1D function : Computation of the updated IMSE for two candidate points

2.5 Uncertainty studies

Once accurate predictors are available for the output properties of interest, they can be used instead of the simulator to investigate the dynamic behavior of a large number of models within the chosen uncertainty range. In particular, they can be used in uncertainty studies to propagate the input uncertainty on the outputs using a Monte Carlo approach. A large sample of the parameter space is generated following the parameter probability

laws, and the corresponding values of the output properties are estimated from the meta-models. This provides an estimation of the outputs distribution corresponding to the uncertainty chosen on the inputs. Analyses can then be conducted on these distributions, for instance with the computation of percentiles in each grid blocks. Note however that the resulting spatial distribution for a given percentile does not necessarily correspond to the values simulated for one of the points of the parameter space. Another analysis consists in estimating in each grid block the probability of occurrence of some geological events, such as the probability for the sediment thickness or sand thickness to lie within a given interval, or the probability that both properties lie simultaneously within these intervals [10]. This uncertainty propagation thus provides an overview of the prior uncertainty on the outputs, that can be used for instance to check the validity of the chosen uncertainty with regards to the available data. It can also be updated if additional information on the parameters become available, for instance using sediment thickness maps derived from seismic data as illustrated in section 3.4.

Quantitative sensitivity studies can also be envisioned to assess the impact of the parameter uncertainties on each output of interest. For instance, Sobol' indices can be estimated based on the meta-model predictions [30]. They quantify the part of the output variance resulting from the parameter uncertainty, considering the effect of the parameter alone (main effect) or the combined effect with other parameters (interactions). These indices range between 0 and 1 and get closer to 1 when the part of the output variance explained by the parameter increases. The global sensitivity of the output to a parameter can be estimated by the total effect, equal to the sum of all the Sobol' indices involving this parameter [16]. For spatial properties, Sobol' indices can be estimated in each grid block, providing an overview of the distribution of the parameters influence in the basin and highlighting regional disparities (see [10] for instance).

3. Application to stratigraphic forward modeling

The approach proposed above is now assessed for risk analysis in stratigraphic forward modeling. We refer to the Dionisos model [11, 12, 14] to simulate the evolution of sedimentary basin stratigraphy and morphology over geological time as described in section 3.1, considering two case studies derived from real data sets and introduced in section 3.2. The first one represents a clastic environment and the second one a carbonate platform. The experiments conducted to assess the potential of the proposed sequential workflow are described in section 3.3, and the results presented in section 3.4. Finally, uncertainty analyses based on the resulting predictors are proposed to illustrate the applicability of the approach.

3.1 Stratigraphic forward modeling

Dionisos is a process-based stratigraphic forward model which aims to characterize the sedimentary basin infill, and in particular the thickness and sedimentary facies spatial distributions of each layer. All sedimentary facies present in the basin are assumed to be described by a finite number of sediments, such as sand and clay for clastic sedimentary systems, and carbonate grains, stromatolites or mud for carbonate sedimentary systems. A stratigraphic simulation is performed in a forward way, from the past up to the present, in a sequence of time steps. At each time step, three main groups of parameters are required, that define respectively the accommodation space (e.g., basin deformation induced by subsidence; sea and lake level changes ...), the sediment supply (e.g., basement and sediment erosion; carbonate production ...) and the sediment transport.

At first, the initial accommodation space and the creation or removal of accommodation space during the simulated time interval are defined by a set of maps. The sea or lake level changes are then specified by an elevation vs. age curve.

Second, the supply of water and sediments is defined along the boundaries of the simulated domain through the characterization of a few sources (width, location and discharge). The production of carbonate sediments is defined by production rate vs. water depth laws.

Third, the transport of each sediment in the basin is calculated using a large-scale law containing two terms: the first one corresponds to slow slope-driven creeping acting mainly along the flanks of carbonate platforms, and the second one to faster water- and slope-driven transport of clastic sediments from the deltas to the open sea.

The creeping used can be expressed as:

$$Q_{s,i} = -(K_{c,i} + K_{w,i} q_w^m \|\nabla h\|^{n-1}) \nabla h$$

where $Q_{s,i}$ [$\text{m}^2 \cdot \text{y}^{-1}$] is the flux for the sediment i , $K_{c,i}$ [$\text{m}^2 \cdot \text{y}^{-1}$] is the slow slope-driven creeping coefficient for this sediment, $K_{w,i}$ [$\text{m}^2 \cdot \text{y}^{-1}$] is the fast water- and slope-driven coefficient, q_w is the dimensionless water discharge flowing over the land and seafloor of the basin, h [-] is the gradient of the seafloor elevation, and m and n are two constants.

3.2 Test cases

Case 1 – Clastic environment

The first case study describes a clastic environment and is derived from the geology and data of the Colville Basin in the National Petroleum Reserve of Alaska [e.g., 17, 24]. This basin is part of the foreland basin lying to the north of the Brooks Range in Alaska. This study is focused on the late Cretaceous Colville Group, from approximately 89 to 65 Ma and covers an area of 240km X 120km. The initial bathymetry of the basin is given in Figure 6(a). It exhibits a continental plateau on the western part, and some east-west convexity in the marine part. The long-term accommodation is defined by the subsidence map between 89 and 65 Ma given in Figure 6(b). It is due to the thrust and folding of the Brooks Range [1] and appears more important in the south-western part. An uplift is also visible in the north-west, corresponding to the forebulge of the foreland basin. The resulting bottom topography at the end of the simulation period is given in dark blue in Figure 6(c). These maps were estimated from the interpretation of seismic data and correlation of wireline logs [4, 5, 17]. The eustasy curve, which drives the short-term accommodation variations, is defined as a combination of the long and short-term Haq curves [15]. Sediments consisting of shale and sand enter the basin through water input on the western margin of the domain (indicated by the blue arrow in Figure 6(a)) and are transported according to water-driven and gravity-driven linear diffusional processes.

For simulation, the domain is discretized horizontally into 120 X 60 grid blocks of size 2km X 2km. The simulated output grid consists of 240 sedimentary layers, each one corresponding to a 0.1 My period.

The uncertainty study focuses on 9 input parameters specified in Table 1 and characterizing the sediment source, sediment transport and eustasy. These parameters are assumed to follow uniform distributions within the chosen intervals. The uncertainty on the eustasy curve is represented by a stretching coefficient that perturbs the initial reference curve along the sea-level axis as illustrated in Figure 7. The study then aims to quantify and analyze the impact of the uncertainty chosen for these parameters on the sedimentary basin infill, and more specifically on the horizontal distribution, at the end of the simulation period, of the total deposited sediment thickness, concentration and thickness for sand and shale in each column of the grid. The mean and standard deviation of sediment thickness, shale thickness and sand proportion for a LHS sample of 40 models are given in Figure 8. No deposition occurs in the grid blocks plotted in white. The largest sediment deposition occurs on average in the southern part of the marine environment, with a predominance of shale. Sand is mainly deposited near the source and the continental plateau. Finally, a strong variability on sediment thickness can be observed in the north-east part of the basin. It can probably be explained by the shape of the basin floor that prevents sediments from reaching this area for a part of the parameter space, in particular for small values of the shale transport coefficient in marine environment.

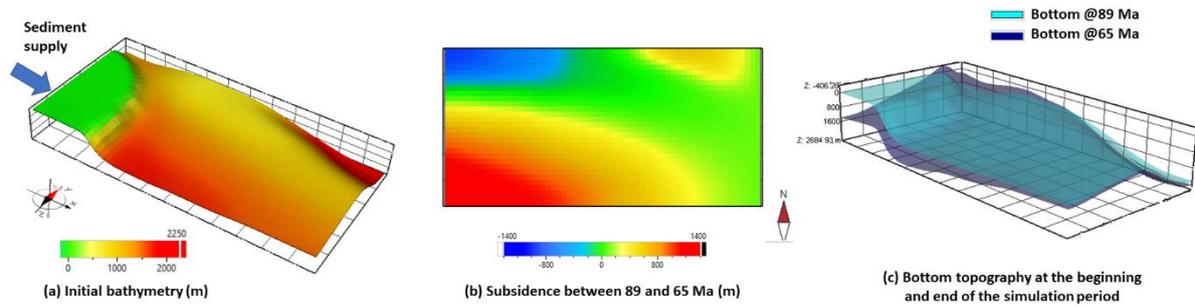


Figure 6 – Case 1, definition of accommodation: (a) initial bathymetry and sediment source location, (b) subsidence map over the simulation period and (c) basin topography at the beginning and end of the simulation period.

Table 1 – Case 1: Ranges of variation chosen for the uncertain input parameters

Parameter		Min	Max
Source definition	Width (km)	65	95
	Sediment supply (km ³ /Ma)	200	5000
	Water discharge (m ³ /s)	450	1350
	Sand concentration (-)	0,1	0,3
Eustasy	Stretching coefficient	0,8	1,2
Transport coefficients	Water-driven – Marine – Sand (km ² /ky)	0,02	2
	Water-driven – Continental – Sand (km ² /ky)	50	400
	Water-driven – Marine – Shale (km ² /ky)	0,2	8
	Water-driven – Continental – Shale (km ² /ky)	500	2500

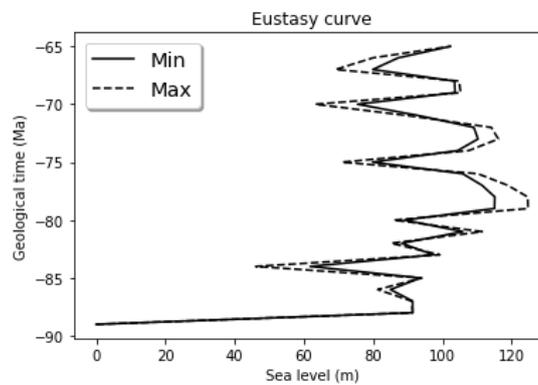


Figure 7 – Case 1: minimal (stretching coefficient = 0.8) and maximal (stretching coefficient = 1.2) variations of the eustasy curve

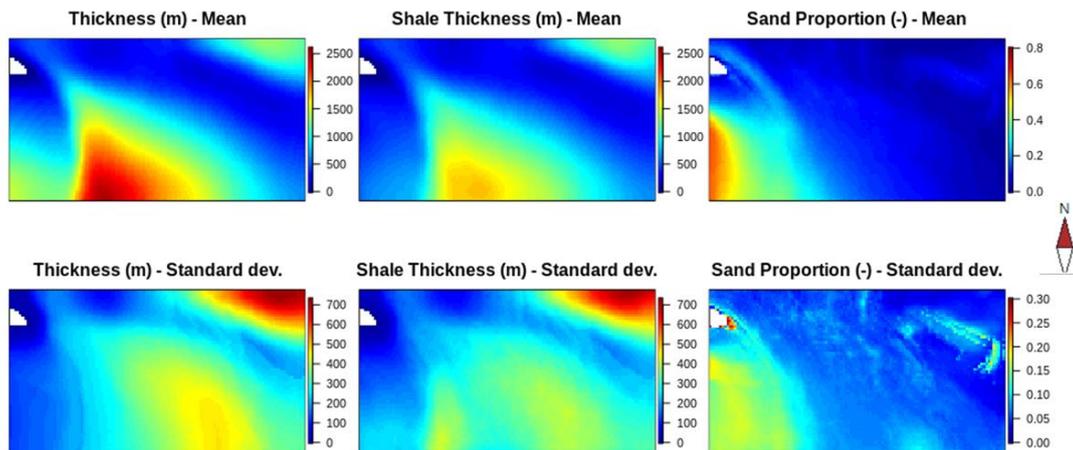


Figure 8 – Case 1 - Mean (first row) and standard deviation (second row) obtained in each column of the grid for a sample of 40 models and three properties of interest (sediment thickness, shale thickness and sand proportion). No deposition occurs in the grid blocks plotted in white.

Case 2 – Carbonate platform

The second case study corresponds to a lacustrine Aptian carbonate platform located in the central portion of the Santos basin [6]. It covers an area of 9km X 34km that is discretized horizontally into 36 X 136 grid blocks of size 250m X 250m for simulation. The basin infill is simulated for a period of 2.4My, with outputs provided every 0.02 My (120 layers in total).

The initial bathymetry of the basin is given in Figure 9(a). It exhibits a north-south elongated continental platform in the center. The long-term accommodation is characterized by a global subsidence. The resulting basin bottom topography at the end of the simulation period is given in dark blue in Figure 9(b). These maps were estimated from interpretation of seismic data and wireline log data [6]. For short-term accommodation variations, the lake level curve is defined from the interpretation of wireline log data in terms of facies and water depth.

For carbonate production, three assemblages are considered: carbonate grains, stromatolites and mud. The production laws considered for these assemblages aim to capture all the biological and physical processes acting at the chosen time and space discretization steps. They are given in Figure 10. Then, a slow slope-driven creeping process is considered for transport [14].

The study conducted here focuses on 12 input parameters related to carbonate production laws, accommodation, and sediment transport. These parameters and their ranges of variations are given in Table 2. They are assumed to follow uniform distributions except for the creeping diffusion coefficients: for these parameters, the uniform distribution applies to their log10 transformation. The corresponding variations of eustasy, carbonate production curves and initial bathymetry are illustrated in Figure 11. For carbonate production laws, two parameters are considered to stretch the initial curves along the depth and production axes: the maximum production rate, and the depth corresponding to this maximum value. For eustasy, a stretching is applied to the initial curve in the sea level direction similarly to Case 1. Finally, a reference map for initial bathymetry is perturbed from its upper and lower points: the bathymetry at these points is changed proportionally to a stretching coefficient and the resulting values are propagated to the whole map by interpolation.

Seven output spatial properties are considered in the study to assess the potential of the proposed sequential workflow. They correspond to the horizontal distribution, at the end of the simulation period, of the total deposited sediment thickness, concentration and thickness of each carbonate assemblage (carbonate grains, stromatolites and mud) in each column of the grid. The mean and standard deviation of the total sediment thickness and carbonate proportions for an LHS sample of 40 models are given in Figure 12. The carbonate grains and stromatolites assemblages are mainly deposited on and near the initial continental platform, with higher

proportions for carbonate grains. Mud is the main sediment type deposited in the distal marine environment, where the variability in deposited sediment thickness is also the largest.

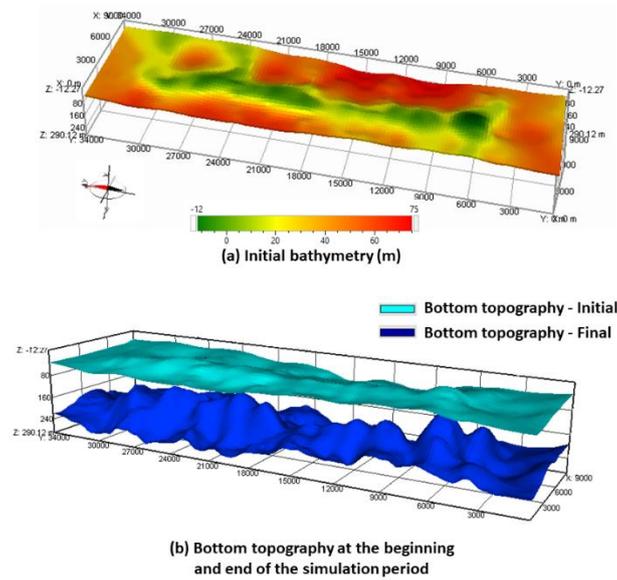


Figure 9 – Case 2, definition of accommodation: (a) initial bathymetry and (b) final and initial basin topography characterizing subsidence over the simulation period

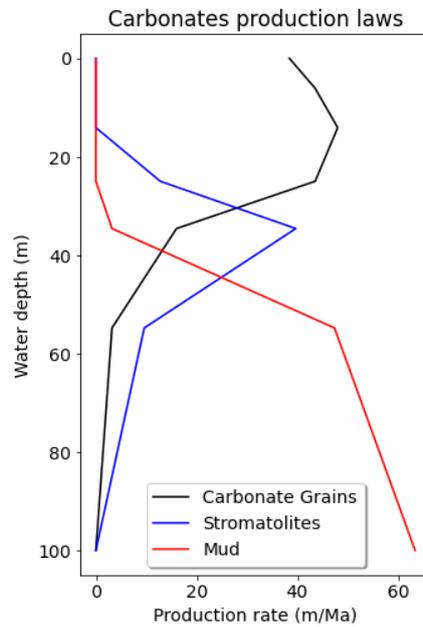


Figure 10 – Case 2: production laws for the three carbonate assemblages as functions of water depth

Table 2 - Case 2: Ranges of variation chosen for the uncertain input parameters

Parameter		Min	Max
Carbonate production laws	Carbonate Grains – Depth of max. prod. rate (m)	7.05	21.15
	Carbonate Grains – Maximum production rate (m/My)	32	96
	Stromatolites – Depth of max. prod. rate (m)	17.3	51.9
	Stromatolites – Maximum production rate (m/My)	32	96

	Mud – Depth of max. prod. rate (m)	50	150
	Mud – Maximum production rate (m/My)	32	96
Eustasy	Stretching coefficient (-)	0,5	1,5
Creeping diff. coef. in marine env. (km ² /ky) $K_{C,i}$	Carbonate Grains	0.35	0.7
	Stromatolites	0.35	0.7
	Mud	0.35	0.7
Bathymetry	Stretching coefficient applied to the initial map lower point	0.8	1.2
	Stretching coefficient applied to the initial map upper point	0.8	1.2

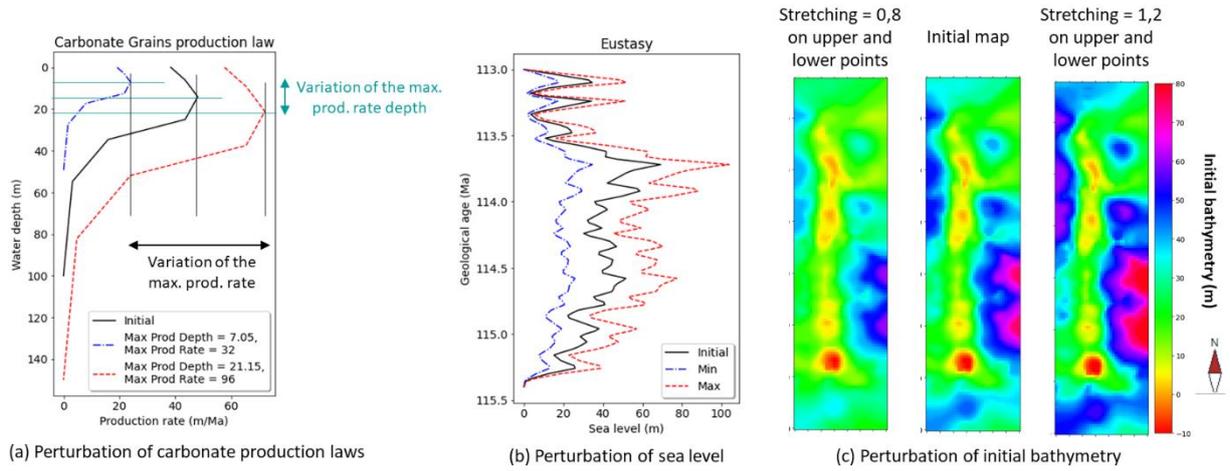


Figure 11 – Case 2: impact of parameter perturbations on the production laws, eustasy curve and initial bathymetry

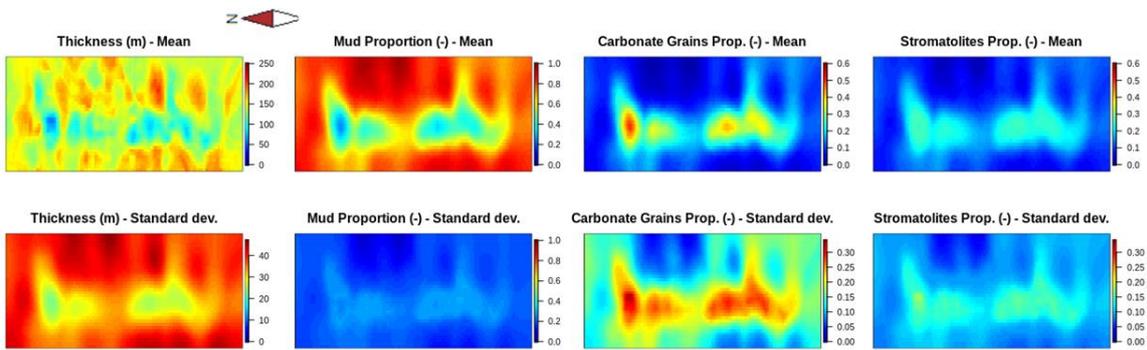


Figure 12 – Case 2: Mean (first row) and standard deviation (second row) obtained in each column of the grid for a sample of 40 models and four properties of interest (sediment thickness, mud, carbonate grains and stromatolites proportions)

3.3 Sequential experiments

To apply the sequential workflow described in section 2 to the two test cases, we consider first each output property of interest in separate processes. In addition, as described in section 2.2, the POD and meta-modeling are applied to the square root of sand thickness and concentration for test case 1, and to the square root of the carbonate grains and stromatolites properties for test case 2. Using this transformation globally provides here better results, probably because the considered properties exhibit many low values in the basin. An initial sample of the parameter space is generated using Latin Hypercube Sampling (LHS). It is of size 10 for the first test case, and of size 15 for the second one, so with a few more simulations than the number of uncertain parameters. For each output property, this sample is complemented sequentially using the three criteria described above and

considering 1 and 5 points added per iteration for each criterion. In total, 6 different sequential methods are thus evaluated per output. As meta-models can vary from one training set to the other, each of these experiments is repeated from four other initial LHS of the same size to obtain more robust comparisons and conclusions (these starting samples being the same for all experiments and properties of a given test case). The components retained in the POD decomposition correspond to a percentage of cumulated variance of 98% for test case 1 and 99% for test case 2. Finally, all experiments are stopped after 150 simulations in total.

This ensemble of experiments is complemented with non-sequential ones that serve as a basis for comparison: LHS of various sizes up to 150 are considered to build predictors for the outputs of interest. As previously, five different LHS of a given size are generated to obtain more robust results.

3.4 Results

The first part of this section is dedicated to the presentation and analysis of the results in terms of efficiency and prediction accuracy. The second part describes uncertainty analyses performed with the resulting estimators.

Meta-modeling

To estimate the accuracy of the predictors obtained through the sequential processes, we introduce for both test cases an additional independent LHS of size 40, the test ensemble \mathbf{D}_{test} , and we perform the corresponding simulations. The R^2 coefficient (2) computed for this ensemble serves in the following to compare the various experiments in terms of efficiency.

Figure 13 and Figure 14 show the evolution, for each output property, of the median R^2_m value (3) averaged over the 5 similar experiments as a function of the number of simulations constituting the training set. The blue curves correspond to the MMSE criterion, the red curves to the IMSE criterion and the green ones to local MMSE. The initial samples are complemented here with 1 point per iteration ($q = 1$ in Algorithm 2). These results are compared to the non-sequential ones obtained with LHS of various sizes generated independently (from 10 or 15 up to 150) and given by black circles. We can observe that, in all cases, the predictor accuracy increases rapidly with the number of simulations at the beginning of the process before a much slower improvement phase: for a sufficient number of simulations in the training set, the new ones added to the ensemble only make it possible to slightly improve predictions. For the first test case (Figure 13), the MMSE and L-MMSE criteria appear the less efficient with regards to the R^2_m value and can provide a slightly lower accuracy on average than LHS-based training sets of the same size. On the contrary, the IMSE criterion appears globally equivalent to LHS. For the second test case (Figure 14), the local MMSE criterion (L-MMSE) can lead to a better efficiency compared to global MMSE at the beginning of the process (carbonate grains and stromatolites properties), but then its performances increase more slowly and can become worse at the end of the experiments. In addition, the current implementation of this approach is quite long as the identification of points in the target Voronoi cell may be time consuming. Compared to LHS, the MMSE and IMSE criteria can provide better results for some properties, in particular for the carbonate grains and stromatolites assemblages. Globally, the IMSE seems here the best trade-off : it provides better results than LHS for the properties that are the most difficult to predict (stromatolites properties in test case 2) and globally equivalent results otherwise.

The evolution of the R^2 coefficient spatial distribution during the sequential experiments is illustrated in Figure 15 for sand proportion in test case 1, and in Figure 16 for carbonate grains proportion in test case 2. These figures show more specifically the spatial distribution of the R^2 coefficient in each grid block averaged over the five similar experiments and obtained at different steps of the process: the LHS-based results are compared to the ones obtained sequentially with the three criteria and 1 point added per iteration. Grid blocks in white correspond to negative R^2 values. These results complement the ones obtained previously with the R^2_m index. Indeed, we can see a global strong improvement of the predictions between the initial samples and the training sets of size 50 in test case 1, and of size 60 in test case 2. At that point, the R^2 value is close to one in many grid blocks, what limits the possible improvement during the following iterations. For test case 1, the IMSE criterion leads to the fastest improvement as already highlighted previously, while the L-MMSE criterion appears the less

efficient after 90 simulations. In addition, the sand concentration seems more difficult to predict accurately in terms of R^2 value in the north part of the basin. This area corresponds however to very low concentration values (see also Figure 8), and more simulations could be considered to improve the predictions if the region is of interest for further analysis. However, the trends are globally reproduced as will be illustrated later on. For test case 2, Figure 16 highlights the fastest accuracy improvement at the beginning of the process using the L-MMSE criterion. However, the results become equivalent for all approaches after some time as the proportions are well predicted in all grid blocks.

The impact of the number of points added to the sample per iteration is summarized in Figure 17 and Figure 18. These graphs show the evolution of the average median coefficient R^2_m as a function of the number of simulations in the training set, considering for each criterion 1 point (solid lines) and 5 points (lines with circles) added per iteration. Here, the simulations are assumed to be performed sequentially to enable comparisons in terms of accuracy. The results highlight that the batch version of the algorithm ($q > 1$ in Algorithm 2) does not degrade the results, even if the covariance properties are not updated between the identification of two successive points within each iteration. As a result, the identification of several points per iteration makes it possible to gain time without loss of efficiency if the corresponding simulations can be run simultaneously, for instance on a computer cluster.

The results discussed so far refer to the application of the sequential workflow to a single spatial output at a time. In practice, the uncertainty study may concern several output properties, identified at the beginning of the study or afterwards for additional analyses. As explained in section 2.4, several options can be envisioned to integrate multiple outputs simultaneously in the sequential process. Here, we rather assess the ability of a sampling identified sequentially for a given property to provide accurate estimations for other outputs of interest. Figure 19 and Figure 20 show, for each output property, the evolution of the average R^2_m values for the predictors of this property built from all the sequential training sets based on the IMSE criterion applied to the property itself and to the other ones. Here, five points are added per iteration. We can observe that, for a given property, the accuracy obtained with the sequential samples based on the other outputs may be lower. However, some properties seem to provide a good trade-off, leading to satisfactory results for all other properties. This is the case for instance for sand proportion in test case 1, and for carbonate grains proportion in test case 2. The global level of accuracy reached at the end of the sequential process with these two properties is illustrated in Figure 21 and Figure 22. They compare, for two models of the test samples, the values simulated for each output properties with the ones predicted using the training set obtained at the end of one of the IMSE-based experiments applied to sand proportion for test case 1 and to carbonate grains proportion for test case 2. Differences can be noticed, but the trends are globally reproduced. These results could probably be improved with additional points in the training set.

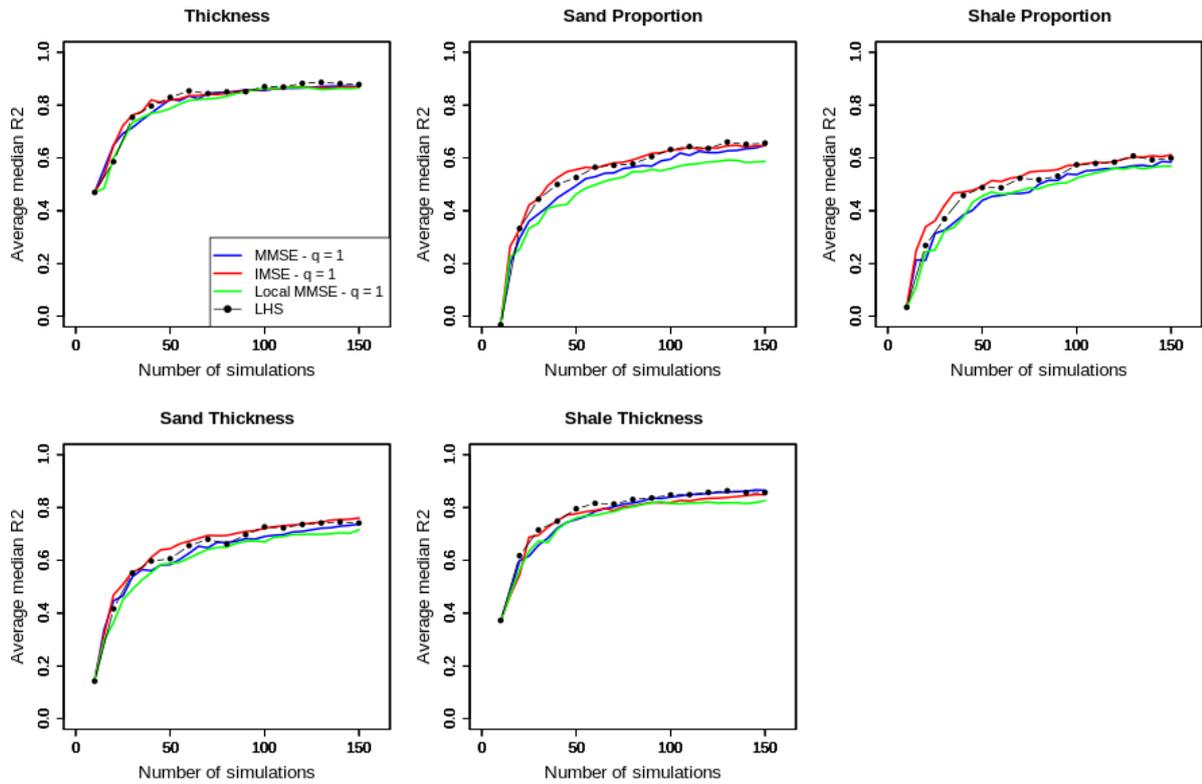


Figure 13 – Test case 1: for each output property of interest, evolution of the median value of the R2 coefficient in the grid blocks ($R2_m$) averaged over the five similar experiments as a function of the number of simulations in the training set

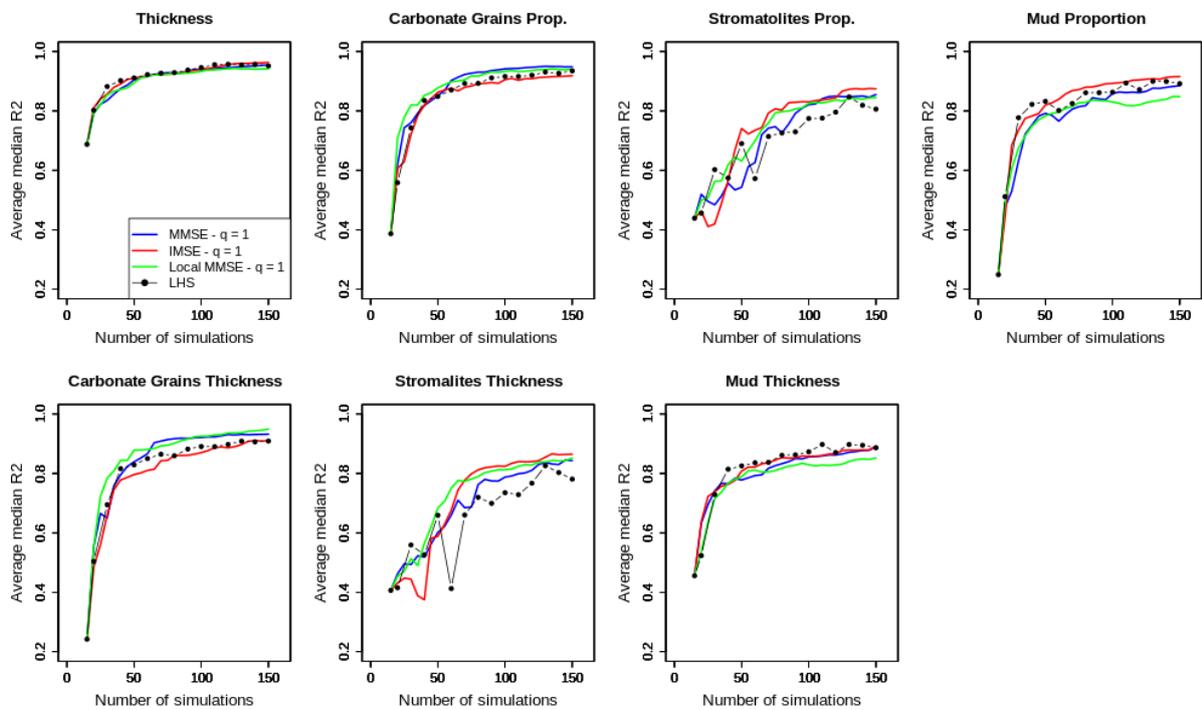


Figure 14 – Test case 2: for each output property of interest, evolution of the median value of the R2 coefficient in the grid blocks ($R2_m$) averaged over the five similar experiments as a function of the number of simulations in the training set

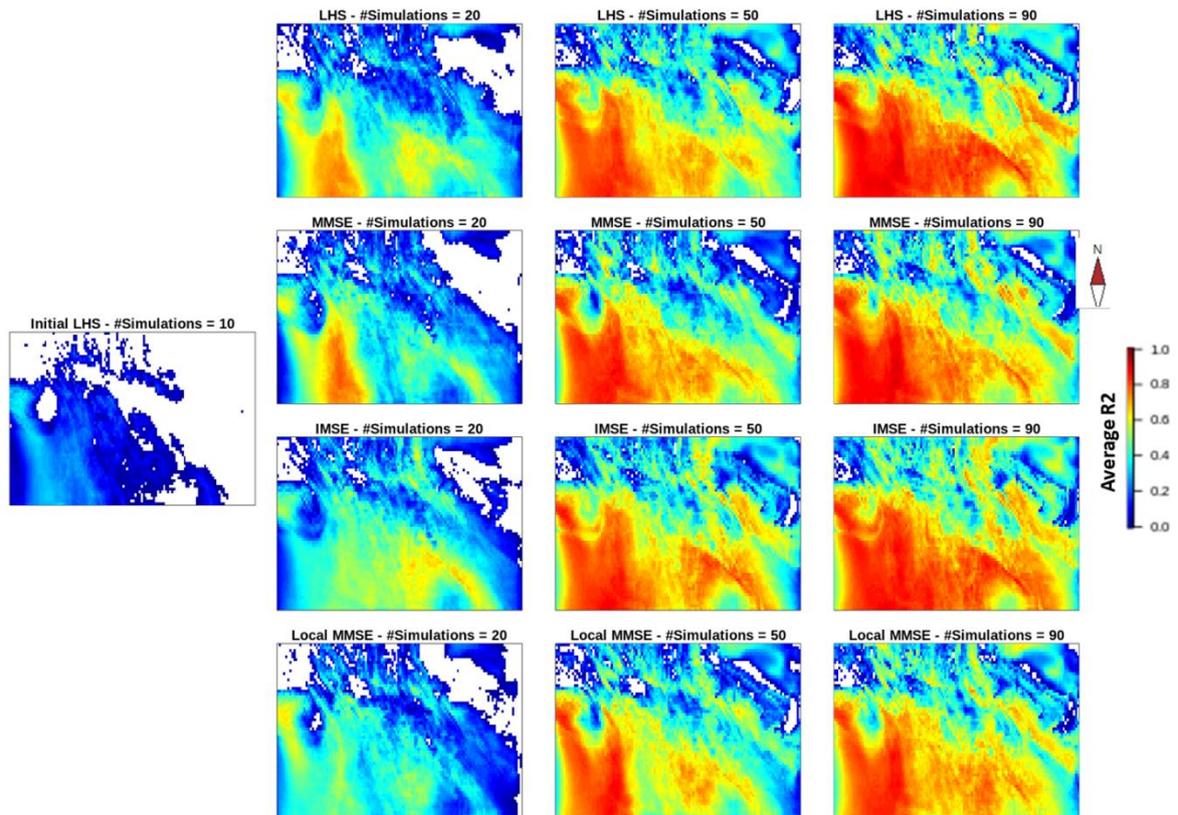


Figure 15 – Test case 1, sand proportion: evolution of the R2 coefficient in each grid block averaged over the five similar experiments, for the initial samples (first column) and samples of size 20 (column 2), 50 (column 3) and 90 (column 4), corresponding to LHS (first line) or obtained during sequential processes with the MMSE criterion (line 2), the IMSE criterion (line 3) and the local MMSE criterion (line 4).

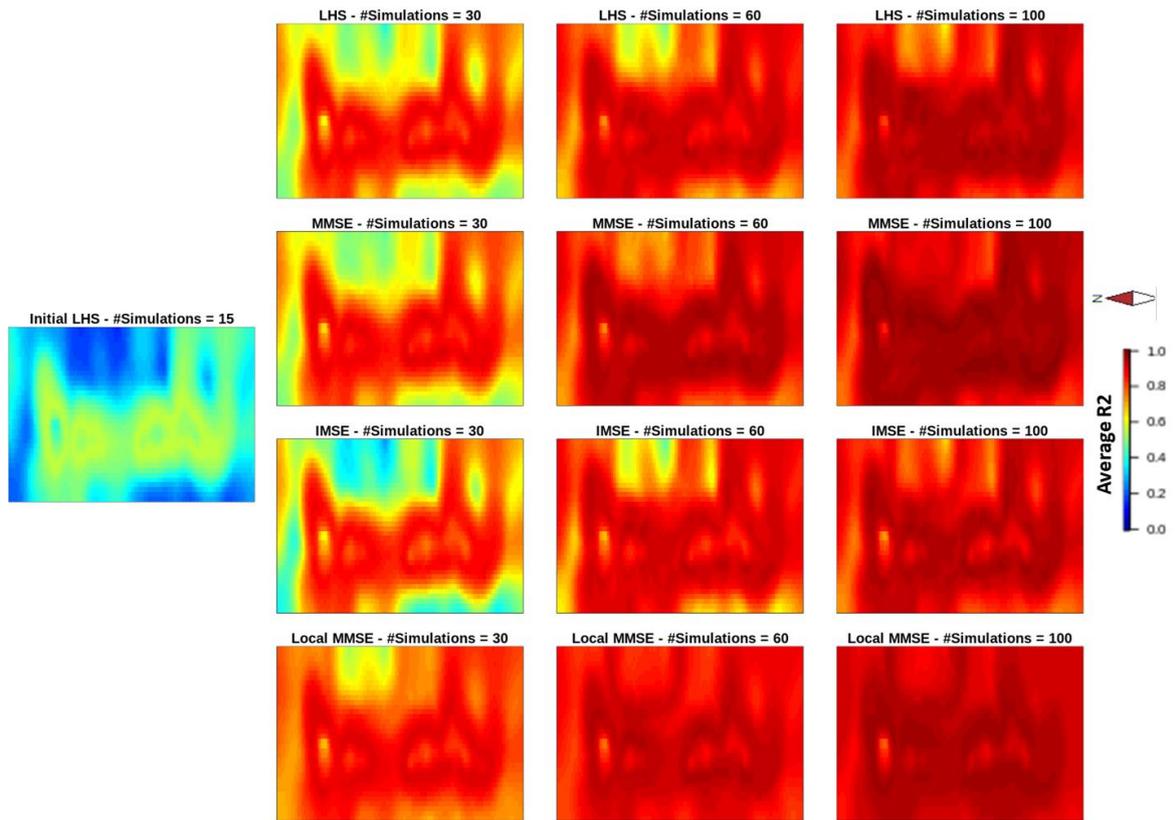


Figure 16 – Test case 2, Carbonate Grains proportion: evolution of the R2 coefficient in each grid block averaged over the five similar experiments, for the initial samples (first column) and samples of size 30 (column 2), 60 (column 3) and 100 (column 4), corresponding to LHS (first line) or obtained during sequential processes with the MMSE criterion (line 2), the IMSE criterion (line 3) and the local MMSE criterion (line 4).

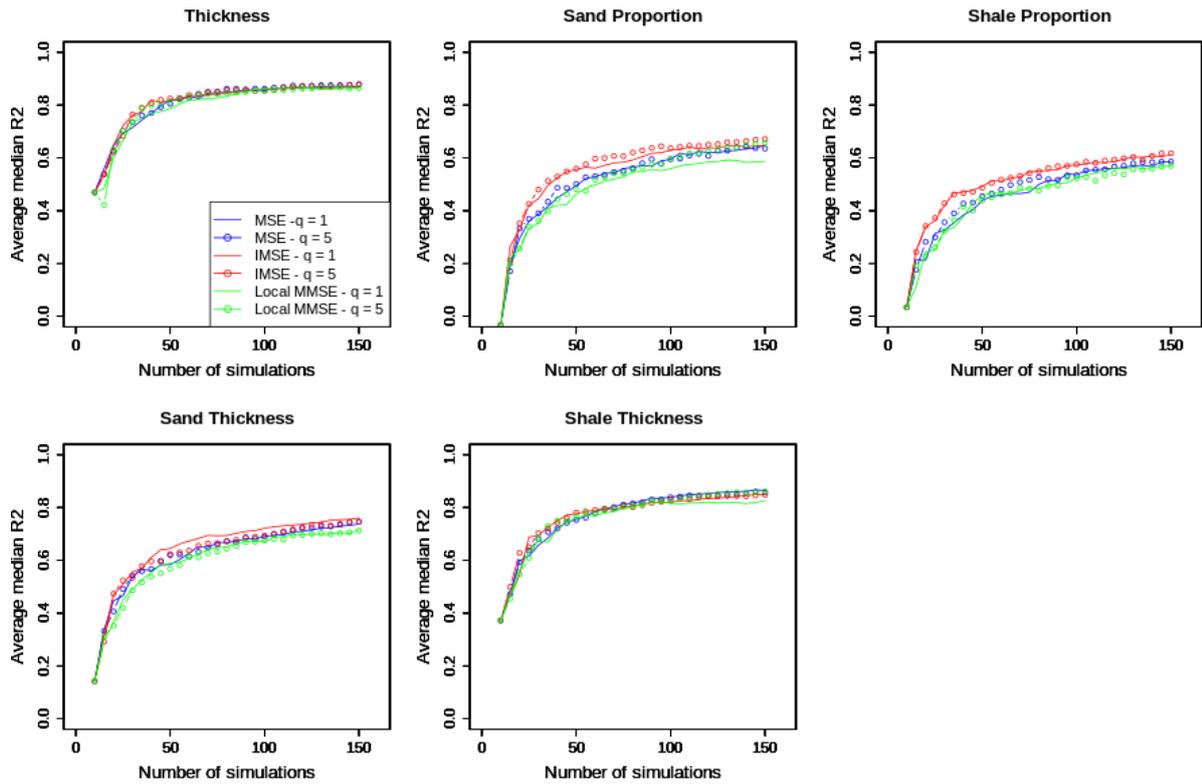


Figure 17 – Test case 1: for each output property of interest, evolution of the median value of the R2 coefficient in the grid blocks ($R2_m$) averaged over the five similar experiments as a function of the number of simulations in the training set, considering 1 point (solid lines) and 5 points (lines with circles) added per iteration. All simulations are assumed to be performed sequentially.

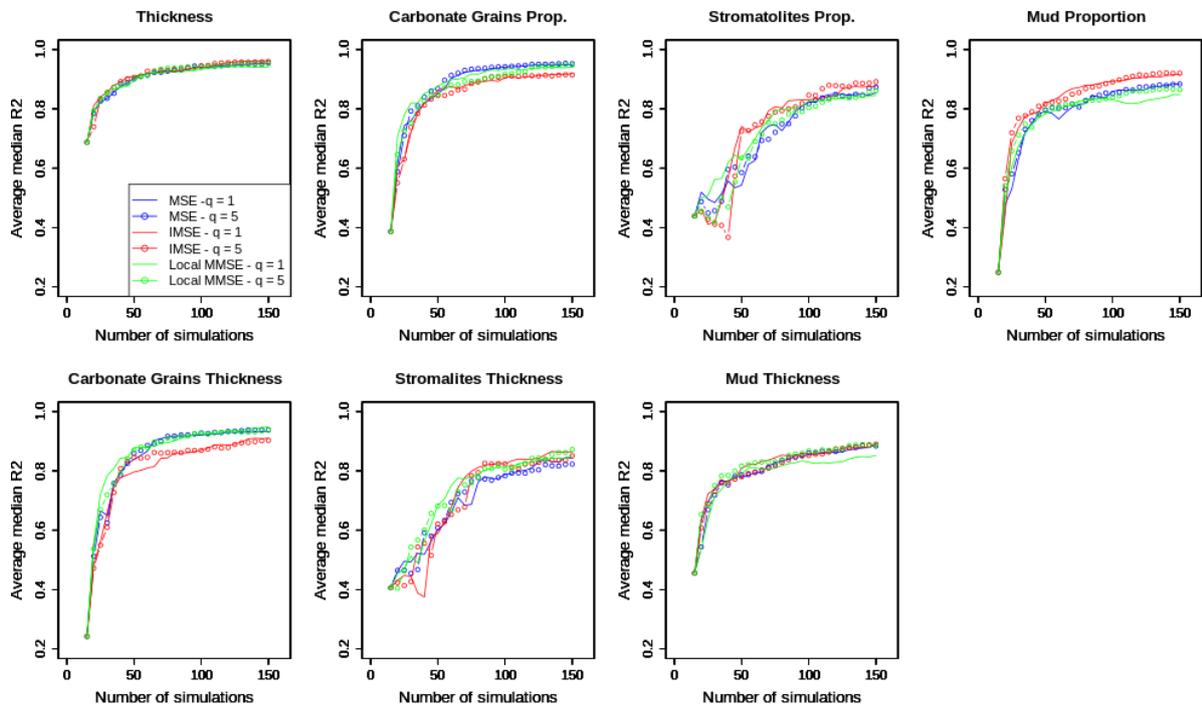


Figure 18 – Test case 2: for each output property of interest, evolution of the median value of the R2 coefficient in the grid blocks ($R2_m$) averaged over the five similar experiments as a function of the number of simulations in the training set, considering 1 point (solid lines) and 5 points (lines with circles) added per iteration. All simulations are assumed to be performed sequentially.

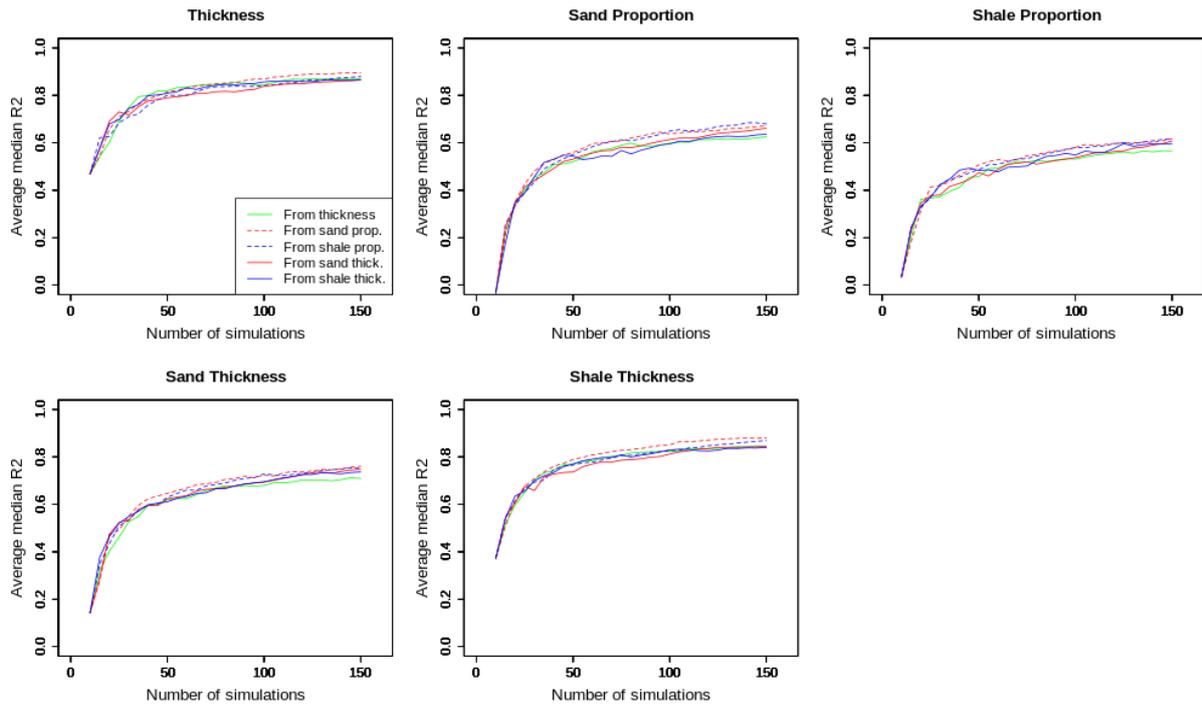


Figure 19 – Test case 1: for each property of interest, evolution of the average median value of the R2 coefficient in the grid blocks ($R2_m$) using the samples obtained with the IMSE criterion and 5 points per iteration applied to all properties

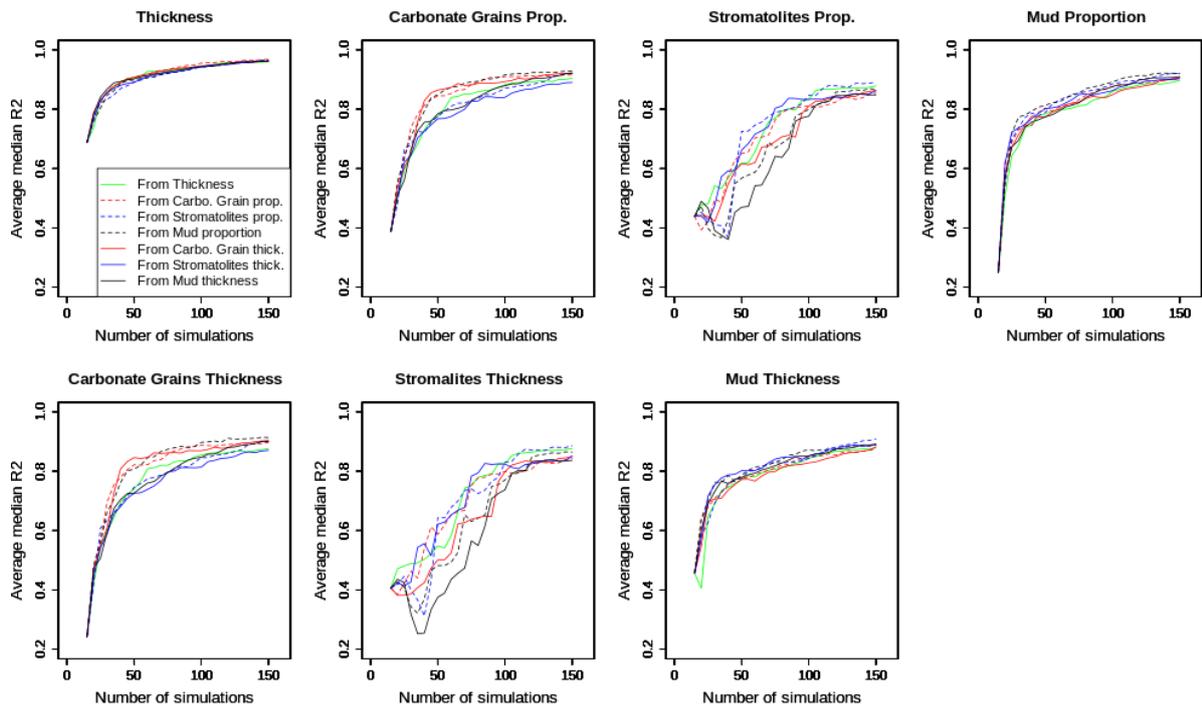


Figure 20 – Test case 2: for each property of interest, evolution of the average median value of the R2 coefficient in the grid blocks ($R2_m$) using the samples obtained with the IMSE criterion and 5 points per iteration applied to all properties

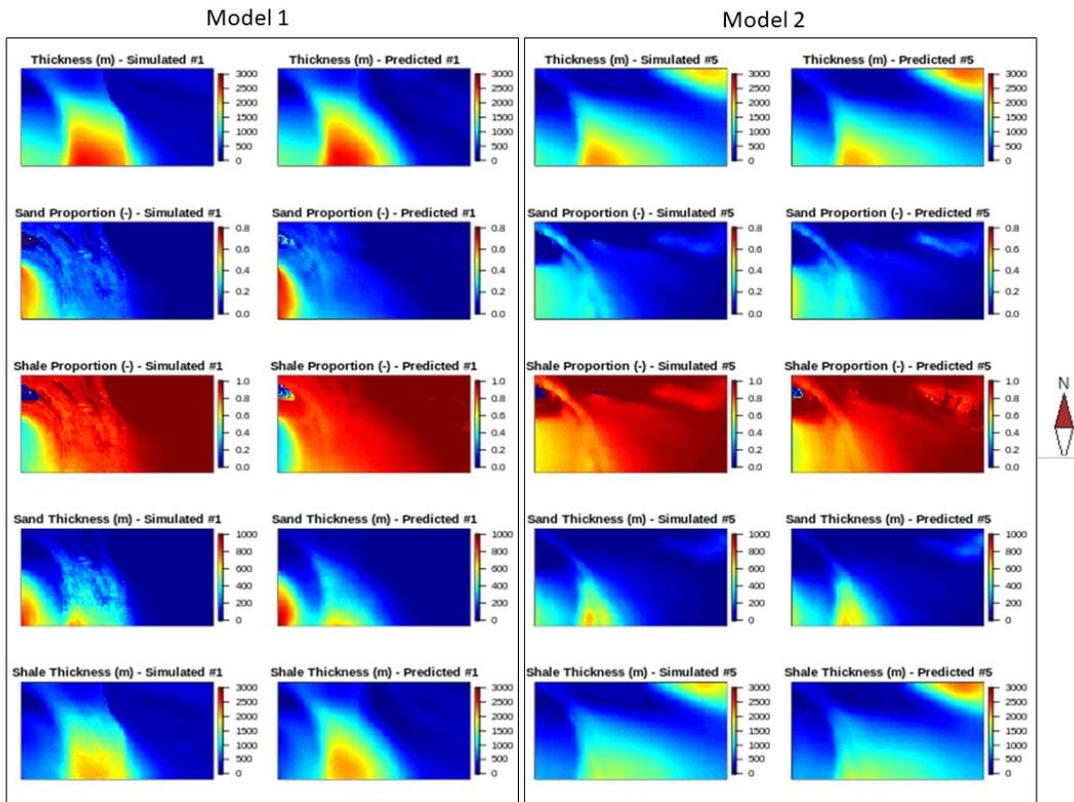


Figure 21 – Test case 1: Simulated and predicted values of the properties of interest for two models of the test sample, using the training set obtained at the end of one of the IMSE-based sequential workflows applied to sand proportion

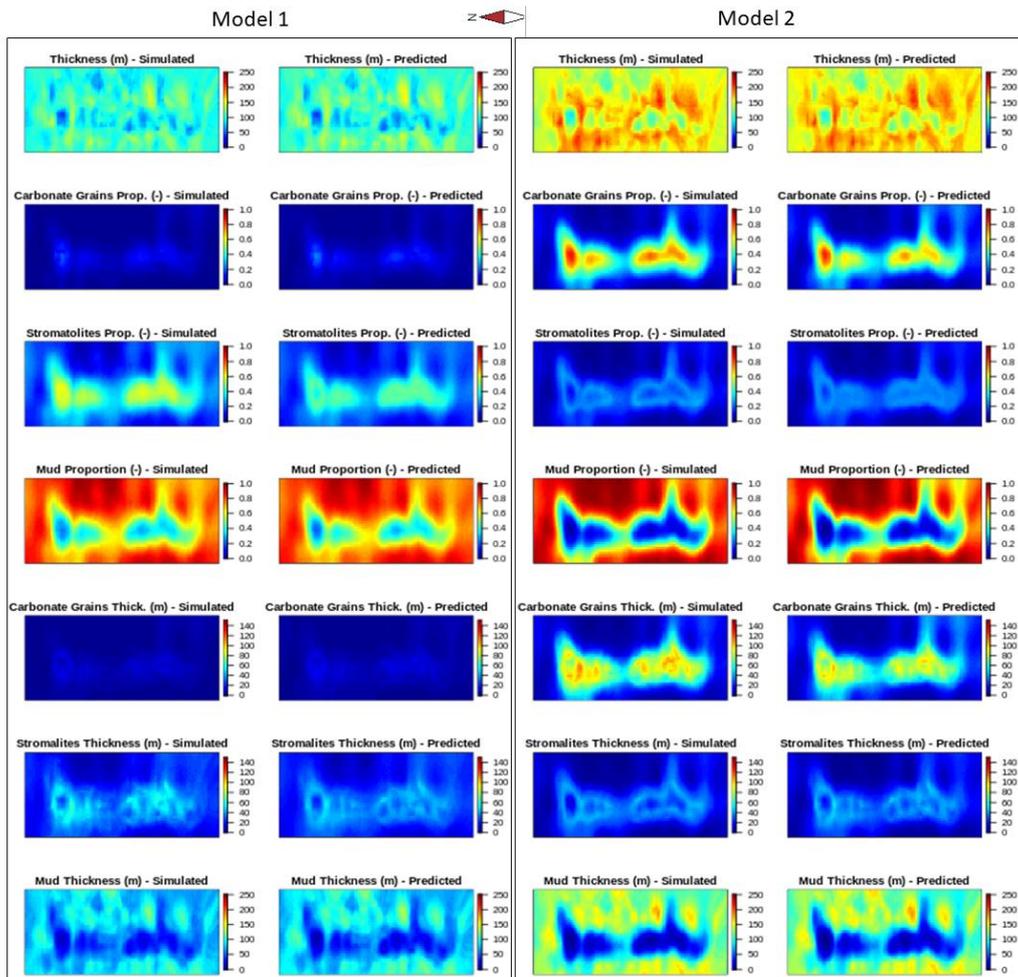


Figure 22 – Test case 2: Simulated and predicted values of the properties of interest for two models of the test sample, using the training set obtained at the end of one of the IMSE-based sequential workflows applied to carbonate grains proportion

Uncertainty analysis

Once the predictions obtained for the outputs of interest are satisfactory, they can be integrated in the uncertainty study.

Case 1 - Clastic environment

For the first test case, we consider the predictors built on the training set of size 150, obtained with the IMSE criterion applied to sand proportion with 5 points added per iteration and illustrated in Figure 21. The P10, P50 and P90 percentile maps obtained for sand thickness are given in Figure 23, top row. Note that these percentiles are computed independently in each column so that the percentile map does not necessarily correspond to the values simulated for a given model. As shown in Figure 24, the distribution of the total deposited sediment thickness is mainly influenced by three parameters for the chosen input uncertainty: water discharge, sediment supply and shale diffusion coefficient in the marine environment. More specifically, the sediment deposition is mainly driven by the value of the sediment supply, except in the north-east part of the basin. As already mentioned in section 3.2, this area is located on the other side of a convexity in the initial topography. The sediment thickness in this area then mainly depends on the value of the shale diffusion coefficient and water discharge that condition the ability of the sediments to pass over the convexity.

Let us now assume that a data map y_{ref} is available for the sediment thickness, derived for instance from seismic acquisition. In this study, we simply choose a reference model within the uncertainty space that is not used in the sequential process and from which we extract the sediment thickness map given in Figure 25(a). Let us now illustrate how we can exploit the current sample to gain information on the parameters from these data and help

in the calibration process. We introduce a quantification of the error E between the reference data map and any other thickness map y (that can be simulated or estimated from meta-models) using the least-square formulation: $E(\theta) = \frac{1}{2N} \sum_{i=1..N} (y(\theta, x_i) - y_{ref}(x_i))^2$ for any set of parameter values $\theta \in \Omega$. The value of this error computed for the 150 models of the training set is given by the red points in Figure 26 as a function of each input parameter value. The black circles correspond to the error predicted for another sample of size 4000 using the kriging-based meta-model built on \sqrt{E} with the 150 models and exhibiting a satisfying R^2 coefficient of 0.96. These results suggest that we could envision to reduce the variation interval on the three parameters that mainly impact the sediment thickness, namely water discharge, sediment supply and shale transport coefficient in marine environment. This uncertainty propagation also makes it possible to identify new models with potentially low error values that could be used as starting points for calibration. Finally, it can be considered to constrain the uncertainty on the other outputs of interest as a first approximation. For instance, the percentiles on the sand thickness map computed from the subsample exhibiting an error lower than 4000 is given in Figure 23 (second row). The P50 values appear closer to the reference sand thickness map given in Figure 25(b). Note however that this subsample does not correspond to the parameter posterior distribution conditioned to the data. In addition, for real cases, sampling criteria dedicated to optimization such as the Expected Improvement [18] should be considered prior to this analysis to identify additional regions of the parameter with low error values. This is however beyond the scope of our study.

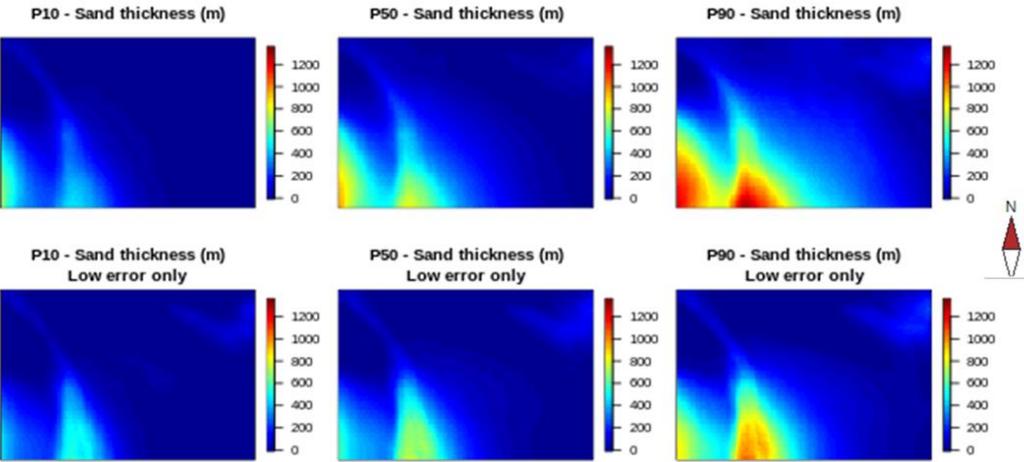


Figure 23 – Test case 1: P10, P50 and P90 percentiles computed on sand thickness for the complete Monte Carlo sample (top) and for the subsample corresponding to predicted errors lower than 4000 (bottom)

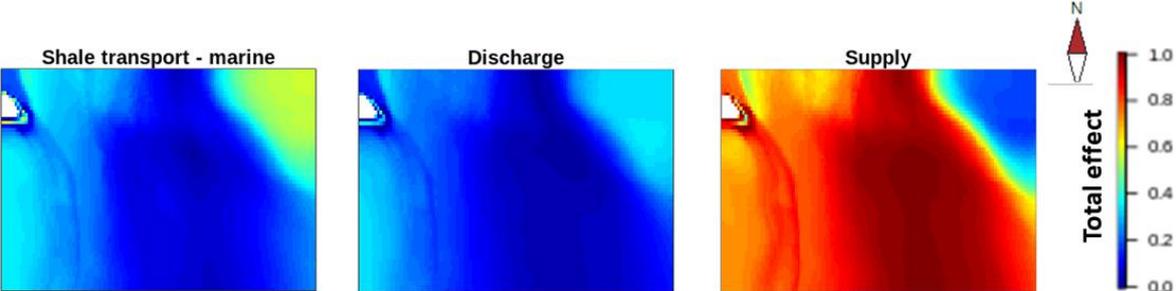


Figure 24 – Test case 1: Total effect computed for the three parameters that mainly impact the deposited sediment thickness

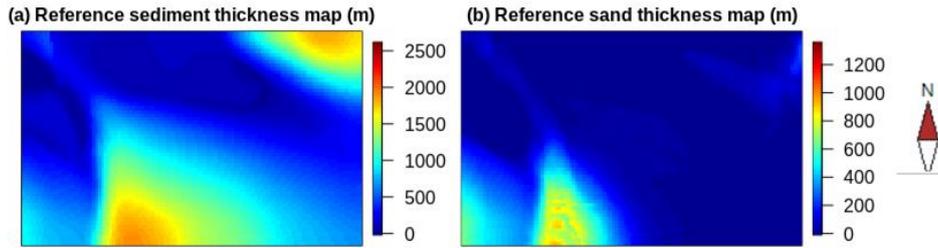


Figure 25 – Test case 1: reference sediment and sand thickness maps

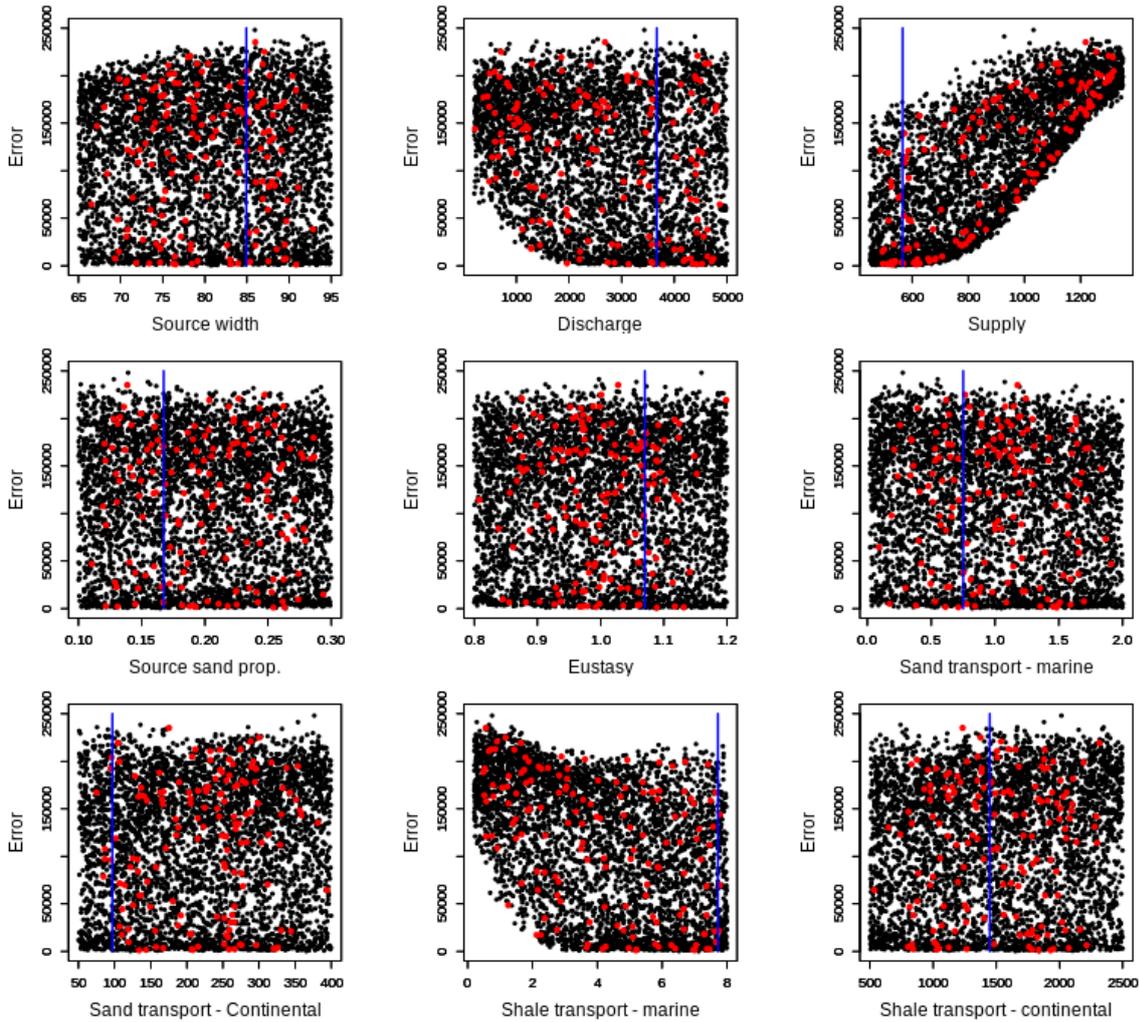


Figure 26 – Test case 1: Error value computed for the training set (red circles) and predicted for the Monte Carlo sample (black circles) as a function of each input parameter value. The blue vertical lines indicate the parameter values used to generate the data.

Case 2 - Carbonate platform

Let us now focus on risk analysis for the carbonate case. As previously, sensitivity indices and percentile maps can be computed for the outputs. Here, we rather focus on a new property of interest, the reservoir thickness. This geological object can be approximated as the carbonate sediment thickness in each column of the grid. Following [13], we rather define it in a diachronous way: in each column of the grid, it is taken equal to the cumulated thickness of all the cells located below the lowest sedimentary deposit containing more than 10% mud, with some tolerance on the bottom layers. The mean and standard deviation of the reservoir thickness for

the test sample of 40 models are given in Figure 27(a)(b). This thickness is thus the largest on and near the initial continental platform, with a very small variability in the other areas. The evolution of the prediction accuracy for this property during the sequential process based on the IMSE criterion applied to carbonate grains proportion with 5 points added per iteration (illustrated in Figure 22) is given in Figure 27(c) as a function of the sample size (on the left of the vertical bar). The distribution of the average $R2$ value in each grid block at the end of this process (sample of 150 models) is given in Figure 27(d). The reservoir thickness thus appears satisfactorily predicted in most parts of the basin, except on the boarder of the thickest area (blue cells). As a result, we started a new sequential process following Algorithm 1 from the training set of 150 models, applying the IMSE criterion directly on the reservoir thickness. The resulting evolution of the median coefficient $R2_m$ is given on the right of the vertical bar in Figure 27(c). The sequential process thus enables to improve further the predictions as also illustrated in Figure 27(e).

The resulting predictions can then be considered to compute maps of percentiles, to estimate the probability for the reservoir thickness to be larger than a given value (Figure 28), or to perform a sensitivity study (Figure 29). We could also envision, as previously, to use these predictions as a starting point for calibration. Finally, Figure 30 summarizes the analysis performed on the total reservoir volume deduced from the reservoir thickness maps. The predictor built for this scalar property with the training set appears very accurate as illustrated in Figure 30(b). The Sobol' indices provide results that are consistent with the ones obtained for the reservoir thickness map (Figure 30(a)) while uncertainty propagation makes it possible to estimate the probability distribution for the reservoir volume (Figure 30(c)).

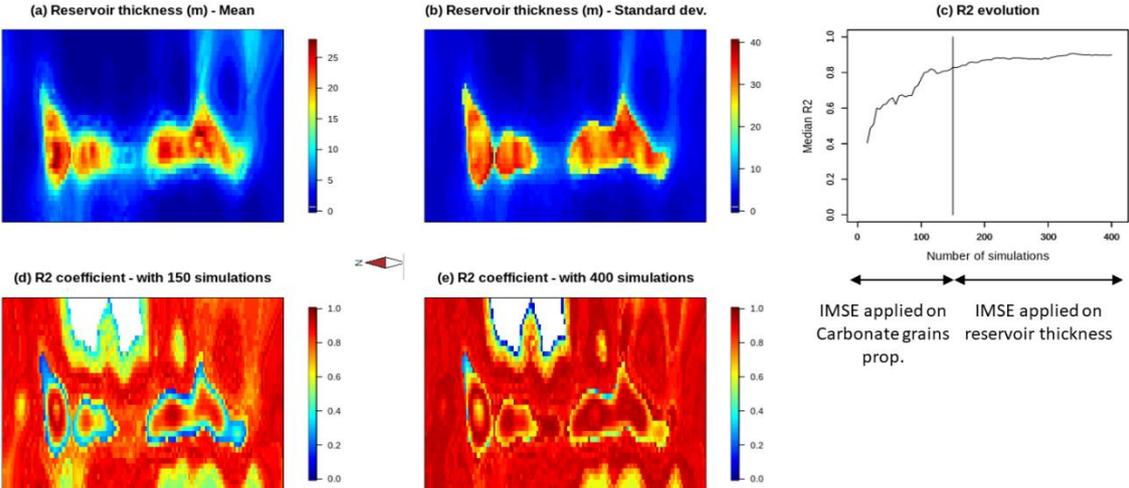


Figure 27 – Test case 2, reservoir thickness: (a) mean values and (b) standard deviation for the test sample, (c) evolution of the average median value $R2_m$; (d) spatial distribution of the $R2$ value in the basin with a training set of 150 and (e) 400 simulations

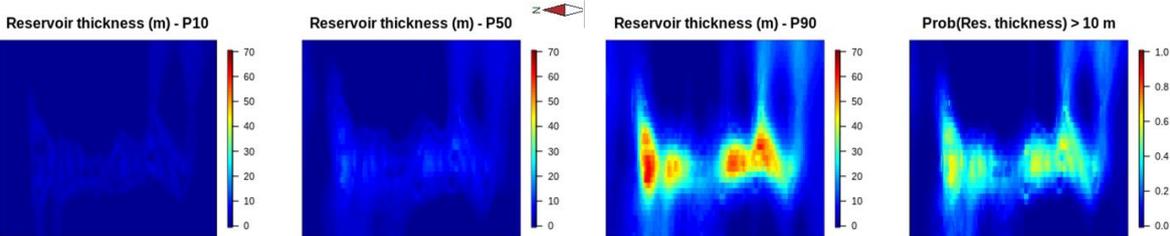


Figure 28 – Test case 2: Reservoir thickness percentiles and probability maps

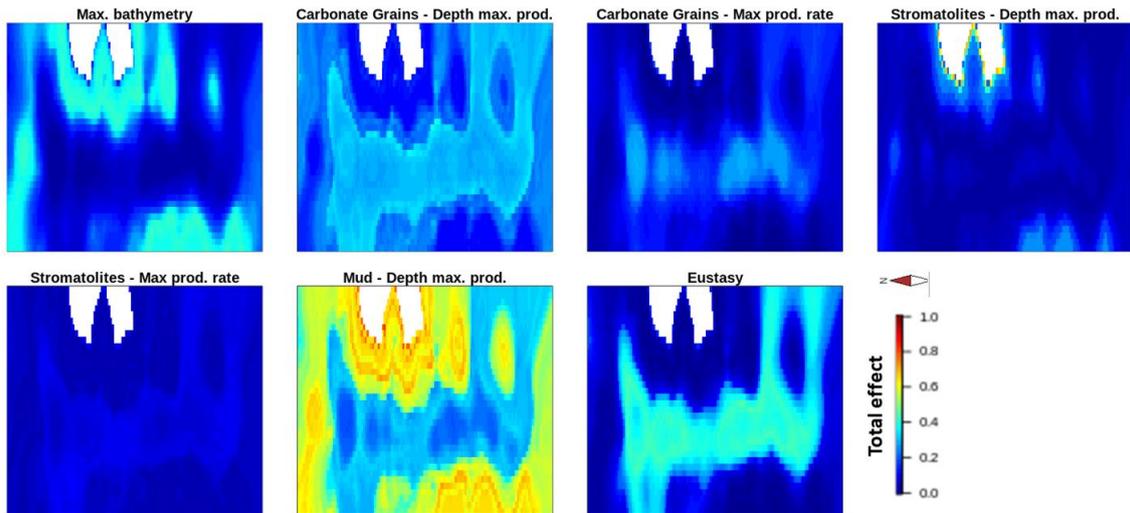


Figure 29 – Test case 2: Total effect computed for the seven parameters that mainly impact the reservoir thickness

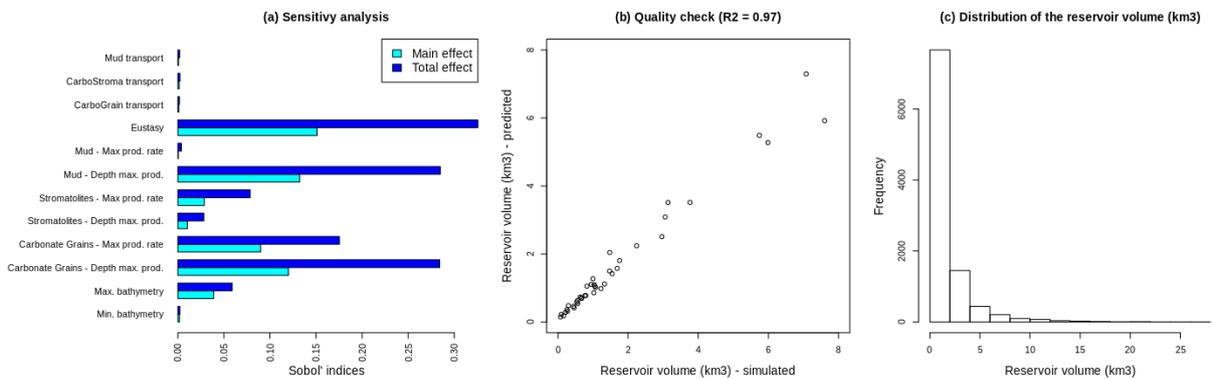


Figure 30 - Test case 2 : (a) main and total effects estimated for the total reservoir volume, (b) simulated vs predicted value of the reservoir volume for the test set and (c) distribution of the reservoir volume for the uncertainty space

4. Conclusions and perspectives

In this paper, we developed a sequential workflow that automatically complements an initial training set to improve kriging-based predictions for spatial outputs of interest. This approach is generic and can be combined to multiple adaptive criteria. Here, three of them were considered for validation on two synthetic stratigraphic forward models, corresponding to a clastic environment and a carbonate platform: the MMSE criterion, the local MMSE and the IMSE. The results on various output properties show that the approach makes it possible to improve the accuracy of the predictions. No sequential criterion clearly outperforms the others here. This may be due to the fact that they all rely on the use of the kriging variance as an uncertainty indicator on the predictions, or by the shape of the input-output relationship to be approximated. However, the IMSE appears globally the most robust in our experiments: it outperforms here LHS-based training sets for the output properties that are the most difficult to predict and provides equivalent results otherwise. The proposed workflow thus makes it possible to iteratively build training sets with a satisfactory efficiency in terms of simulation time and prediction accuracy. The user does not need to bother about the training set definition, what should facilitate the applicability of the kriging-based risk analysis workflow. In addition, several models can be added at a time to the sample without degrading the efficiency of the process, and the approach can be applied for multiple properties simultaneously or sequentially, for instance if a new output of interest comes up during

the study. However, the use of adaptive criteria does not clearly improve the process efficiency here compared to LHS for many properties. This may be due to the fact that a globally homogeneous sample of the parameter space is required to correctly approximate these outputs. Some properties still require a non-negligible number of models to be properly estimated. It could thus be interesting in future work to evaluate other iterative criteria, e.g., geometrical, or to use adaptive ones differently. Other investigations could concern the reduced-basis decomposition approach or the use of faster models in the training process to reduce simulation times, e.g., defined with coarser discretization in time and space (multi-fidelity meta-modeling).

Finally, uncertainty analyses were conducted with the predictors to highlight the advantages of the approach. In particular, the reasonable time necessary to investigate a large number of models makes it possible to perform statistical analyses, e.g., to quantify the impact of the input parameters and to estimate the output distributions. A preliminary study was also performed for calibration on synthetic data. Even if this work focuses on spatial output properties, meta-models can be built to estimate any other continuous scalar, time-dependent or depth-dependent output, and in particular the cost function that quantifies the error between the observed and simulated data. Depending on the data observed on the basin, it can be envisioned to combine the estimations obtained for various properties in order to estimate the appropriate ones to be compared to the observed data. In future work, it would be interesting in future work to study in more detail the use of meta-models for calibration, e.g., combined to minimization algorithms, and to consider real data for validation.

Acknowledgments

The authors would like to thank the reviewers for their detailed and insightful reviews and very interesting comments and questions that helped improve the manuscript.

V. Gervais and S. Bouquet worked on the methodology development; D. Granjeon designed the case studies; V. Gervais performed the validation tests and analysis. All authors contributed to the writing of the manuscript. They declare they don't have conflict of interest.

Data availability

The synthetic test cases considered in this paper were built from the data provided in [17,24,6].

Appendix

To approximate components $\alpha_k, k = 1, \dots, L$, in the estimator (1), we consider here Gaussian processes (also referred to as kriging-based meta-models): each component is assumed to be a realization of a Gaussian process $A_k(\boldsymbol{\theta})$ defined by

$$A_k(\boldsymbol{\theta}) = m_k(\boldsymbol{\theta}) + Z_k(\boldsymbol{\theta})$$

$\boldsymbol{\theta} \in \mathbb{R}^d$ denotes the input parameters. $m_k(\boldsymbol{\theta})$ represents the mean of the Gaussian process and is chosen constant here ($m_k(\boldsymbol{\theta}) = m_k$). $Z_k(\boldsymbol{\theta})$ is a stationary Gaussian process with zero mean and covariance C_k . It represents the fluctuations around the mean. C_k is defined here with a parametric form as $C_k(\boldsymbol{\theta}, \boldsymbol{\theta}') = \sigma_k^2 R_k(\boldsymbol{\theta}, \boldsymbol{\theta}')$ with $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \mathbb{R}^d$, σ_k^2 being the process variance and R_k the correlation function, chosen as the Matérn 5/2 function [23] in what follows.

Then, the best linear predictor of α_k at a new point $\boldsymbol{\theta}^*$ given $\boldsymbol{\alpha}_k^D$ is:

$$\hat{\alpha}_k(\boldsymbol{\theta}^*) = m_k + \mathbf{r}_k^T(\boldsymbol{\theta}^*) \mathbf{R}_k^{-1} (\boldsymbol{\alpha}_k^D - m_k \mathbf{I})$$

where $\mathbf{I} \in \mathbb{R}^n$ is a vector of ones. $\mathbf{R}_k \in \mathbb{R}^{n \times n}$ and $\mathbf{r}_k(\boldsymbol{\theta}^*) \in \mathbb{R}^n$ are defined as: $(\mathbf{R}_k)_{ij} = R_k(\boldsymbol{\theta}^i, \boldsymbol{\theta}^j)$ and $(\mathbf{r}_k)_j(\boldsymbol{\theta}^*) = R_k(\boldsymbol{\theta}^j, \boldsymbol{\theta}^*)$. The variance of predictor $\hat{\alpha}_k$ at $\boldsymbol{\theta}^*$, also called kriging variance, is given by:

$$\hat{s}_k^2(\boldsymbol{\theta}^*) = \sigma_k^2 \left(1 - \begin{pmatrix} 1 & \mathbf{r}_k^T(\boldsymbol{\theta}^*) \end{pmatrix} \begin{pmatrix} 0 & \mathbf{I}^T \\ \mathbf{I} & \mathbf{R}_k \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ \mathbf{r}_k(\boldsymbol{\theta}^*) \end{pmatrix} \right)$$

It represents the model mean square error and equals zero at the sampling points \mathbf{D} .

The quality of estimator $\hat{\alpha}_k$ can be assessed using cross-validation as described in section 2.3 (Equation (4)), or using the test ensemble $\mathbf{D}_{test} = \{\tilde{\theta}^1, \dots, \tilde{\theta}^{n_{test}}\}, \tilde{\theta}^i \in \Omega$. In this case, its predictivity can be assessed as in Equation (2) by:

$$R2(\hat{\alpha}_k) = 1 - \frac{\sum_{j=1}^{n_{test}} (\alpha_k(\tilde{\theta}^j) - \hat{\alpha}_k(\tilde{\theta}^j))^2}{\sum_{j=1}^{n_{test}} (\alpha_k(\tilde{\theta}^j) - \bar{\alpha}_k)^2}$$

For all $j \in \{1, \dots, n_{test}\}, \alpha_k(\tilde{\theta}^j)$ is the projection coefficient of $y(\tilde{\theta}^j, \mathbf{X})$ on the basis vector $\phi_k(\mathbf{X})$, and $\bar{\alpha}_k$ the mean of the resulting n_{test} coefficients.

Bibliography

1. Blythe, A. E., Bird, J. M., Omar, G. I. (1996) Deformational history of the Central Brooks Range, Alaska: Results from fission-track and 40Ar/39Ar analyses. *Tectonics*, 15(2), 440–455. doi: 10.1029/95TC03053
2. Bruneau, B., Villié, M., Ducros, M., Chauveau, B., Baudin, F., Moretti, I. (2018). 3D Numerical Modelling and Sensitivity Analysis of the Processes Controlling Organic Matter Distribution and Heterogeneity—A Case Study from the Toarcian of the Paris Basin. *Geosciences*, 8(11), 405.
3. Busby, D. (2009) Hierarchical adaptive experimental design for Gaussian process emulators. *Reliability Engineering and System Safety*, 94, pp 1183-1193.
4. Decker, P.L. (2007) Brookian Sequence Stratigraphic Correlations, Umait Field to Milne Point Field, West-Central North Slope, Alaska. Alaska Division of Geological & Geophysical Surveys, Preliminary Interpretative Report 2007-2, 19pp, 11 sheet.
5. Decker, P.L. (2010) Brookian Sequence Stratigraphic Framework of the Northern Colville Foreland Basin, Central North Slope, Alaska (Poster and Presentation): Dnr Spring Technical Review Meeting, Anchorage, April 21–22, 2010. Alaska Division of Geological & Geophysical Surveys, 30pp, 31 sheet.
6. de Paula Faria, D.L., dos Reis, A.T., de Souza, O.G. Jr. (2017) Three-dimensional stratigraphic-sedimentological forward modeling of an Aptian carbonate reservoir deposited during the sag stage in the Santos basin, Brazil. *Marine and Petroleum Geology* 88, pp. 676-695.
7. Dubrule, O. (1983) Cross validation of kriging in a unique neighborhood. *Journal of the International Association for Mathematical Geology*, 15(6), 687–699. doi: 10.1007/BF01033232
8. Ducros, M., Nader, F.H (2020) Map-based uncertainty analysis for exploration using basin modeling and machine learning techniques applied to the Levant Basin petroleum systems, Eastern Mediterranean. *Marine and Petroleum Geology* 120.
9. Forrester A., Keane, A.J. (2009) Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45, 50-79.
10. Gervais, V., Ducros, M., Granjeon, D. (2018) Probability maps of reservoir presence and sensitivity analysis in stratigraphic forward modeling. *AAPG Bulletin*, 102(4), 613-628.
11. Granjeon, D. (2014) 3D stratigraphic forward model compared with analogue flume model: insights on the non-linear water-driven sediment transport and on the impact of baselevel cycles on continental margin and incised valleys, In: Martinius, A.W., Ravnas, R., Howell, J.A., Steel, R.J., Wonham, J.P. (eds.), *From depositional systems to sedimentary successions on the Norwegian continental margin*. IAS, Special Publication 47, pp. 453-472. ISBN-13: 9781118920466
12. Granjeon, D. (2019) Use of high-performance stratigraphic forward modelling to improve siliciclastic and carbonate reservoir depositional architecture description. *Journal of the Japanese Association for Petroleum Technology*, 84 (1), pp. 59-70.
13. Granjeon D., Gervais V. (2023) Forward and inverse stratigraphic forward modelling in exploration and appraisal workflows: Insights from the Miocene carbonate platforms, Central Luconia Basin, South China. *SEPM Special Publication 114*, 314-333.
14. Granjeon, D., Joseph, P. (1999) Concepts and applications of a 3D multiple lithology, diffusive model in stratigraphic modelling. In: Harbaugh, J.W., Watney, W.L., Rankey, E.C., Slingerland, R., Goldstein, R.H., Franseen, E.K. (eds), *Numerical Experiments in Stratigraphy: Recent Advances in Stratigraphic and*

- Sedimentologic Computer Simulation*, *SEPM Special Publications* 62, pp. 197-210. doi: 10.2110/pec.99.62.0197
15. Haq, B.U., Hardenbol, J., Vail, P.R. (1988) Mesozoic and Cenozoic chronostratigraphy and cycles of sea-level change. *SEPM, Special Publication* 42, pp. 71-108
 16. Homma, T., Saltelli, A. (1996) Importance measures in global sensitivity analysis of model output. *Reliability Engineering & System Safety*, 52(1), 1–17, doi: 10.1016/0951-8320(96)00002-6.
 17. Houseknecht, D.W., Schenk, C.J. (2001) Depositional sequences and facies in the Torok formation, National Petroleum Reserve – Alaska (NPRA). In: Houseknecht, D.W. (ed) Petroleum plays and systems in the National Petroleum Reserve – Alaska (NPRA). *SEPM Sp. Pub.* 21, pp. 179-200
 18. Jones, D., Schonlau, M., Welch, W. (1998) Efficient Global Optimization of Expensive Black-Box Functions. *J. Global Optim.* 13(4), 455–492. doi: 10.1023/A:1008306431147
 19. Johnson, M., Moore, L., Ylvisaker, D. (1990) Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, 26, 131-148.
 20. Le Gratiot, L., Cannamela, C. (2015) Cokriging-based sequential design strategies using fast cross-validation techniques for multi-fidelity computer codes. *Technometrics*, 57(3), pp. 418-427.
 21. Loeve, M. (1978) *Probability Theory*, vol. I–II. Springer, New York.
 22. Marrel, A., Perot, N., Mottet, C. (2015) Development of a surrogate model and sensitivity analysis for spatio-temporal numerical simulators. *Stochastic Environmental Research and Risk Assessment*, 29(3), 959–974. doi: 10.1007/s00477-014-0927-y
 23. Matérn, B. (1986) *Spatial Variation*. Springer, New York
 24. Moore, T.E, Wallace, W.K., Bird, K.J., Karl, S.M., Mull, C.G., Dillon, J.T. (1994) Geology of northern Alaska. In: Plafker, G. and Berg, H.C., (eds.) *The geology of North America*, vol G-1, *The Geology of Alaska*. Geol. Soc. of Amer., Boulder, Colo.
 25. Paola, C. (2000) Quantitative models of sedimentary basin filling. *Sedimentology* 47, Suppl. 1, pp. 121-178
 26. Picheny, V., Ginsbourger, D., Roustant, O., Haftka, R., Kim N-H (2010) Adaptive designs of experiments for accurate approximation of a target region. *Journal of Mechanical Design*, 132(7), pp. 071008-1 – 071008-9.
 27. Roustant, O., Ginsbourger, D., Deville, Y. (2012) DiceKriging, DiceOptim : Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *J. Stat. Softw.* 51(1), 1–55. doi: 10.18637/jss.v051.i01
 28. Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P. (1989). Design and Analysis of Computer Experiments. *Statist. Sci.*, (4), 409–423. doi: 10.1214/ss/1177012413
 29. Santner, T.J., Williams, B.J., Notz, W. (2003) *The design and analysis of computer experiments*. Springer-Verlag, New York
 30. Sobol', I.M. (1990) On sensitivity estimation for nonlinear mathematical models. *Mathematical Modeling & Computational Experiment*, 1 (4), p. 407–414.
 31. Thenon, A., Gervais, V., Le Ravalec, M. (2022) Sequential design strategy for kriging and cokriging-based machine learning in the context of reservoir history-matching. *Computational Geosciences*, 26(5), pp. 1101-1118
 32. Torelli, M., Kowalewski, I., Gervais, V., Wendebourg, J., Dupré, S., Wolf, S., Gout, C., Deville, E. (2021) Quantification of natural microbial methane from generation to emission in the offshore Aquitaine: A basin modelling approach. *Marine and Petroleum Geology*, 127.
 33. Xiao, J., Waltham, D. (2019) Non-uniqueness and symmetry in stratigraphic interpretations: A quantitative approach for determining stratal controls. *Sedimentology* 66, pp. 1700-1715.