# Metamodel-assisted analysis of an integrated model composition: an example using linked surface water – groundwater models.

**Vasileios Christelis* and Andrew G. Hughes**
**British Geological Survey, Keyworth, Nottingham NG12 5GG, UK**
*** corresponding author: vc@bgs.ac.uk**

## Highlights

1) **Distributed groundwater models were linked together with a river model via the OpenMI software platform.**
2) **Kriging metamodels were applied to facilitate analysis with the integrated models.**
3) **The overall computational savings were in the range of 70-90%.**
4) **Monte Carlo simulations demonstrate that metamodels were in good agreement with the responses of the integrated models.**
5) **Sensitivity Analysis using the metamodels accurately identified the important parameters.**

**Keywords: Integrated Modelling; Metamodels; Kriging; OpenMI; Monte Carlo simulation; Sensitivity Analysis**

**Abstract:** Integrated modelling is a promising approach to simulate processes operating within complex environmental systems. It is possible, however, that this integration may lead to computationally expensive compositions. In order to retain the process fidelity without loss of accuracy, the use of Kriging metamodels is proposed to perform Monte Carlo simulation and sensitivity analysis, in lieu of compositions developed using the model linking standard OpenMI. Results from the Monte Carlo simulation showed that the metamodels were in a good agreement with the original responses. However, metamodels provided a less accurate approximation of the original output distribution for the composition which involved a stronger non-linear behaviour. The fast runtimes of the metamodels allowed for increased computational budgets leading to an accurate screening of the important parameters for an Elementary Effects Test. Overall, Kriging metamodels provided significant computational savings without compromising the quality of the outcomes, even using small training data sets.

## 1. INTRODUCTION

The simulation of complex flow and mass transport processes in environmental systems is made possible by running sophisticated numerical codes on desktop computers. These computer models aim to provide a high-fidelity analysis of the various environmental phenomena which often consist of large spatial-scales and long simulation periods. In recent years, there is an increasing interest in environmental systems modelling to simulate the various processes in a holistic view instead of separately running stand-alone models and link their output information in a monomeric manner (Makropoulos et al., 2010; Nalbantis et al., 2011).

Integrated modelling is a relatively recent discipline and aims to facilitate the solution of complex problems by joining models from a range of different sources together (Laniak et al., 2013). At its purest it is taking a model code making it linkable and then allowing the scientist or other end-user to answer their question by joining together models in a bespoke combination (Barkwith et al., 2014). The aim is to develop methods so that questions can be answered that are more complex than can currently be addressed using single instance models. To achieve model integration and depending on the question being posed and the audience for the results, various challenges have to be met, particularly the models themselves need to be:

- Described and made available – metadata catalogues have to exist with access to a way of running the model; this includes documenting the model code development process and the underlying physic / maths and the assumptions made (Harpham and Danovaro, 2015)
- Made linkable – either by using well described file formats or runtime coupling, e.g. OpenMI (Harpham et al., 2016) or CSDMS (Peckham et al., 2013)
- Described properly – e.g. semantics / ontologies, so that different disciplines can recognise the same variable even if its described differently (Nativi et al., 2013)
- Run as quickly as possible – does it need to run in the cloud (High Performance Computing), parallelisation employed or be made available via Web Services (Castronova et al., 2013)?
- Uncertainty properly evaluated for the model chain – e.g. UncertWeb (Bastin et al., 2013)
- Results visualised appropriately, e.g. summaries for decision-makers or full datasets for scientists / higher level users (Voinov et al., 2016).

Despite the promising aspects of the holistic approach for a more accurate representation of the simulated system, there are additional challenges related to the resulting computational burden. Inevitably, the higher the complexity of the combined models and processes, the heavier the computational cost. In that case, the implementation of iterative tasks, such as sensitivity analysis or optimization, is hampered due to impractical computational requirements (Queipo et al., 2005). A competitive approach to handle the computational burden, assuming that the available computer resources rely on a single PC, is to employ fast emulators which mimic the behaviour of the complex physics-based environmental models (Ratto et al., 2012). These approximations are usually referred as surrogate models or metamodels and are constructed using input-output data derived from the original expensive computer simulations (Sacks et al., 1989). For the rest of this work the term metamodel is used. Typically, a set of training patterns is designed, which spans the input variable space as widely as possible, in order to construct a metamodel that preserves a reasonable level of prediction accuracy to unseen data (Forrester and Keane, 2009).

There is a wide body of literature in various science disciplines which considers metamodeling techniques. Artificial Neural Networks (ANN), Polynomial Functions, Kriging, Radial Basis Functions (RBF), Support Vector Machines (SVM) and Gaussian Processes (GP), are popular examples of metamodels which have been used in the environmental literature (Razavi et al., 2012a). ANN appear to be the most common choice in environmental modelling literature with various applications in optimization (e.g. Aly and Peralta, 1999; Yan and Minsker, 2006; Sun et al., 2015; Yazdi et al., 2015; Fienen et al., 2016) in model calibration (e.g. Khu and Werner, 2003; Zou et al., 2009) and uncertainty analysis studies (e.g. Shrestha et al., 2009; Yu et al., 2015). Kriging and RBF metamodels have been mainly preferred in water resources optimization problems (e.g. Bau and Mayer, 2006; Shoemaker et al., 2007; Castelletti et al., 2010; Razavi et al., 2012b; Tsoukalas and Makropoulos, 2015a; Tsoukalas and Makropoulos, 2015b; Christelis and Mantoglou, 2016; Tsoukalas et al., 2016; Christelis et al., 2016; Christelis et al., 2018). They can operate as interpolating metamodels, that is, they pass through all the previously evaluated points with the original model and thus can get more accurate on predicting the response of the original model as new input-output data become available (Forrester et al., 2008). Metamodels with such interpolating capabilities are considered beneficial for deterministic simulation outputs (Wang et al., 2014). Recently, another family of metamodels, the Polynomial Chaos Expansions (PCEs) (Xiu and Karniadakis, 2002), have been applied in environmental modelling to address uncertainty quantification and sensitivity analysis problems (e.g. Rajabi et al., 2015 and references therein;

Babaei et al., 2015; Bellos et al., 2017). Another approach, less common within environmental modelling, is the use of physics-based metamodels which are constructed by combining models of different fidelity (e.g. Bianchi et al., 2015; Christelis and Mantoglou, 2017).

Nevertheless, the selection of the metamodel type, largely depends on the scope of the modelling study and the available information on the underlying behaviour of the physics-based computer model (Broad et al., 2014). In general, metamodeling methods apply satisfactorily given that the original model responses are reasonably smooth at least for the input domain of interest (Forrester and Keane, 2009). In light of this, it becomes apparent that issues such as the range of the input domain, the dimensionality of the input vector or the non-linear input-output relation of the original model, may adversely affect the metamodel accuracy. This can be particularly evident in uncertainty analysis (UA) and sensitivity analysis (SA) studies and constitutes a challenging research field for metamodeling studies (Song et al., 2015). However, SA utilizes simulation-based frameworks which are computationally intensive and the use of metamodels, under certain conditions, may allow for a tractable computational cost (Pianosi et al., 2016). Borgonovo et al. (2012) investigated the usefulness of metamodels in a global SA study for alleviating the computational burden involved in such studies. They found that metamodels successfully reduced the computational cost, however, their accuracy largely depends on the proper metamodel training and the ability of the metamodel to capture the original model structure. Coutts and Yokomizo (2014) used metamodels for global SA and they concluded that even simple metamodels may provide the user with important information about the original model behaviour, however, complex interactions may only be revealed by more sophisticated metamodels which in turn are more complicated and time-consuming. Yu et al. (2014) compared different metamodels for reducing the cost of uncertainty analysis with a flood inundation model. They mentioned the necessity for considering the uncertainty arising from the use of the metamodels in the analysis. Ge et al. (2015) proposed an efficient Kriging-metamodel-based approach for SA for high-dimensional and computationally expensive computer models.

In this work, Kriging metamodels are employed in lieu of a computationally expensive integrated environmental model, in order to conduct Monte Carlo (MC) simulation and SA within reasonable computational times. The original physics-based computer model is a composition of two hydrological models that was built using an implementation of the OpenMI model linking standard and simulates interactions between surface-subsurface model instances. The composition is presented with two variations of the subsurface component, which results in different complexity and runtimes for single composition. This allowed for a broader

assessment of metamodels' efficiency and efficacy. First, the MC simulation is used as a validation set for investigating the impact of metamodel structure and initial training size on predicting the output distribution of the physics-based model. Then, the best identified metamodel structures are utilized for performing a one-at-a-time (OAT) sensitivity analysis run to screen the important input factors of the original model.

## 2. METHODOLOGY

### *2.1 Composition used for study*

The modelling approach consists of two models components in a linked composition (integrated model). The basic groundwater model was built using ZOOMQ3D (Jackson and Spink, 2004) and was configured as a single layer model of 10 km length by 2 km width with a grid mesh of 100 m (see Figure 1). The model has a base of 35 m above Ordinance Datum (aOD, UK equivalent to mean sea level) with a hydraulic conductivity of 5 m/d and 0.5 m/d in two zones (see Figure 1) and uniform recharge of 1.0 mm/d. When the model is operated in confined mode, i.e. transmissivity that doesn't vary with saturated thickness, then the model top of 60 m aOD was used. The model run until steady-state conditions were met, to enable a water balanced solution to be reached and to allow interaction with the river model.

The groundwater model was coupled to a simple river model developed with MCRouter (Mansour et al., 2013). Each node of MCRouter was linked to a corresponding node in the groundwater model. The runtime linkages were undertaken using OpenMI v1.4 (Gregersen et al., 2007) to produce a single composition run using Pipistrelle (Harpham et al., 2014) (see Figure 2). A uni-directional coupling with the groundwater model supplying the river model with baseflow calculated using leakage nodes was used to link the model instances within the composition (Fig. 2) .The model composition was run using the command line version of Pipistrelle to enable it to be controlled by the MATLAB code.

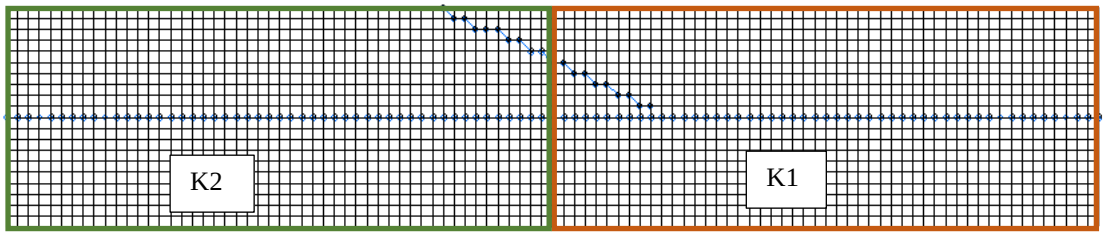


**Fig. 1**. Basic model setup (black gridlines ZOOMQ3D groundwater model; blue dots MCRouter)

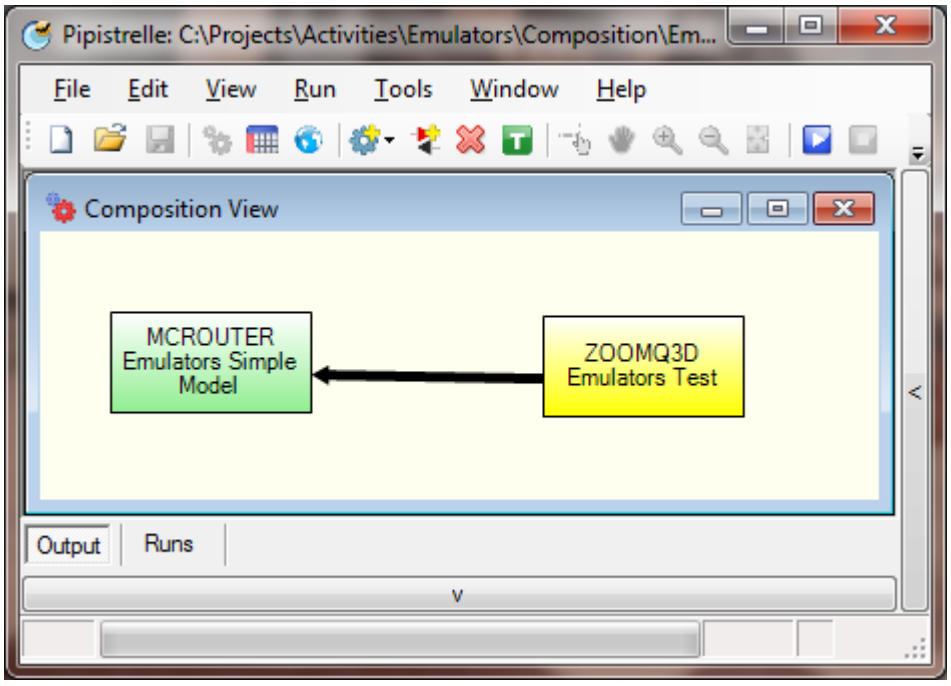


**Fig. 2**. Screenshot of model composition built using Pipistrelle

The two compositions were run until the groundwater heads reach steady-state conditions. As previously discussed two different composition forms have been employed. The composition which comprises the Muskingum-Cunge routing channel model and the confined groundwater flow model, thereinafter called as MCR-GCF, requires an approximate computational time of 4.2 minutes to reach steady-state. On the contrary, the other composition form which comprises the Muskingum-Cunge routing channel model and the unconfined groundwater flow model, thereinafter called as MCR-GUF, reaches steady-state after approximately 63 minutes. The significant difference in the computational burden is due to the increased number of iterations needed for the non-linear unconfined groundwater flow equation to reach steady-state. Examples of the differences in the simulation output between MCR-GCF and MCR-GUF models, as well as, for the time evolution of groundwater heads until steady-state, are presented below (Figure 3 for steady-state heads and Figure 4 for an example time series of groundwater head reaching steady state).

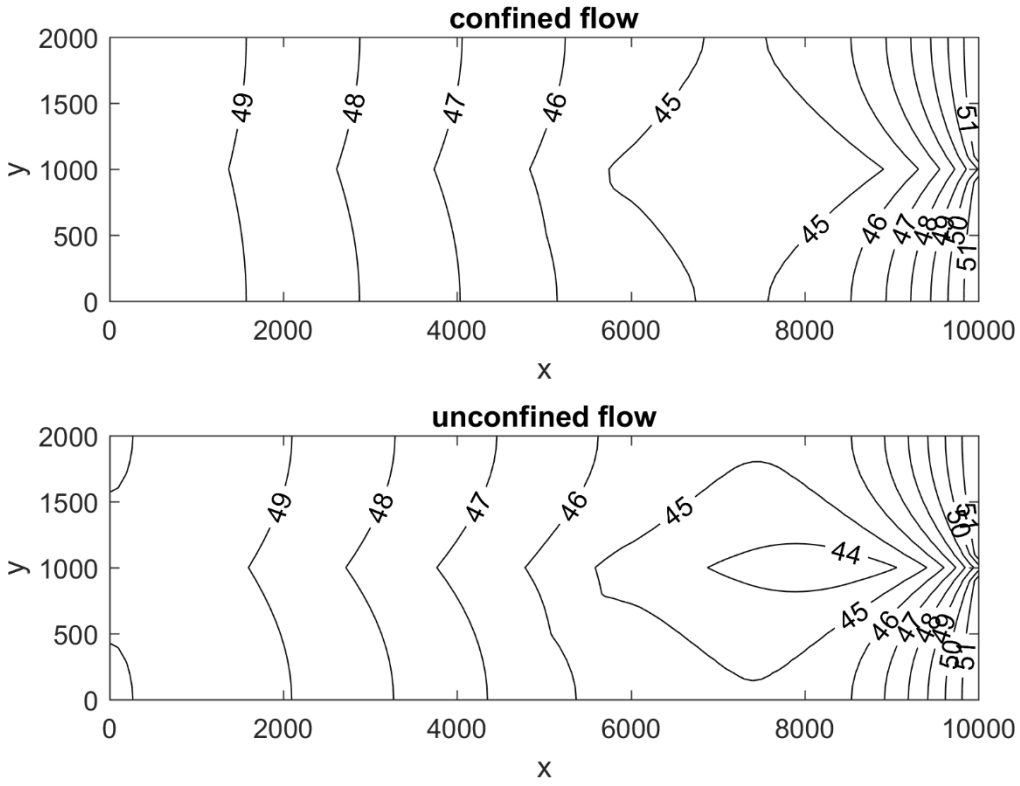


**Fig. 3.** Simulation output of the groundwater flow model component, in the case of confined and unconfined flow for the same input parameter set.

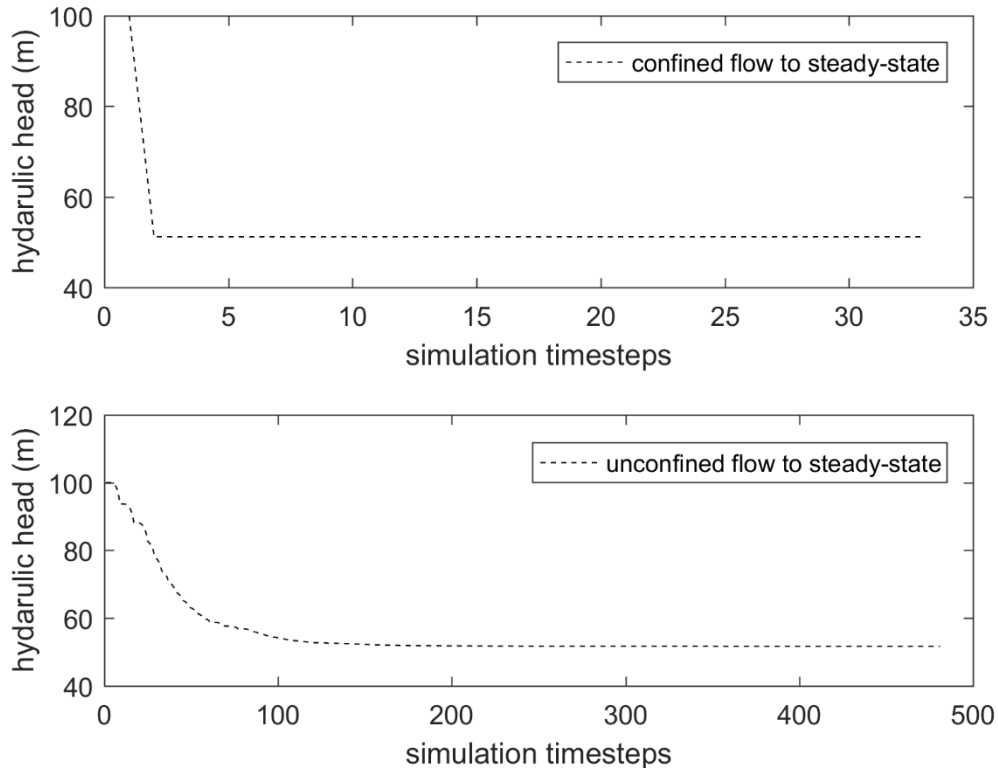


**Fig. 4.** Time evolution of groundwater heads for confined and unconfined flow until steady-state.

*2.2 The metamodel*

The metamodel used in this study is based on the MATLAB toolbox DACE (Design and Analysis of Computer Experiments) (Lophaven et al. 2002). DACE constructs approximation models based on Kriging, a method which was originally developed in the field of geostatistics (Krige 1951). Kriging has been introduced as an emulator to deterministic computer simulations in Sacks et al. (1989). Detailed mathematical background about Kriging metamodeling can be found elsewhere in the literature (e.g. Jones et al. 1998; Lophaven et al. 2002; Forrester et al. 2008; Kleijnen 2009). Here, we only briefly present the general concept of a Kriging emulator as applied in our problem.

Let assume a set of multivariate inputs $\boldsymbol{x}^{(i)}, i = 1, \ldots . n$ where each one includes numerical values of variables, $K1$ , $K2$ and $R_g$ (hydraulic conductivity for the two zones and recharge respectively). A corresponding univariate scalar output $h^{(i)}, i = 1, \ldots . n$, representing hydraulic head, is obtained by running the composition for each input $\boldsymbol{x}^{(i)}$. Kriging treats the deterministic outputs h as if they were generated from a stochastic process where its general trend is described by a regression function $f(\boldsymbol{x})$ and the associated error terms are correlated random variables of a zero mean Gaussian process, with variance $\sigma^2$ and correlation matrix $\boldsymbol{R}$ (Jones et al. 1998). The general concept, which is very convenient for deterministic computer simulations due to the absence of random error, is that if the distance between points $\boldsymbol{x}$ and $\boldsymbol{x}'$ is small, then the approximations $\hat{h}(\boldsymbol{x})$ and $\hat{h}(\boldsymbol{x}')$ of the "true", black-box function (the computer model) should be highly correlated (Forrester et al. 2008). The Kriging model can be expressed as the sum of a regression term $f(\boldsymbol{x})$ and a Gaussian process $Z(\boldsymbol{x})$, defined by the covariance function $Cov\big(Z(\boldsymbol{x}), Z(\boldsymbol{x}')\big) = \sigma^2 R(\boldsymbol{x}, \boldsymbol{x}')$**,** as:

$$\hat{h}(\boldsymbol{x}) = f(\boldsymbol{x}) + Z(\boldsymbol{x}) \tag{1}$$

Different formulations of Kriging (e.g. ordinary, simple, universal) assume either a constant or a polynomial for the regression function $f(\boldsymbol{x})$ while the choice for the correlation function $R(\cdot\,,\,\cdot)$, is often critical for the accuracy of the Kriging approximation model. DACE toolbox allows for regression models of different polynomial order as well as correlation models of different structures. Here, Exponential, Gaussian and Linear correlation models are combined with zero, first and second order regression models. The correlation function involves a set of parameters $\boldsymbol{\theta}$ which are identified using the Maximum Likelihood Estimation method and

numerical optimization techniques (Forrester et al., 2008). This step adds computational cost to the fitting process, particularly as the sample size gets larger. In its original formulation DACE employs a simple, efficient optimization algorithm which can be sensitive to the initial $\boldsymbol{\theta}$ values and might only find local solutions. Here, the lower and upper bounds for $\boldsymbol{\theta}$ values were set to $lb = 0.01$ and $ub = 2$ respectively. The initial guess for $\boldsymbol{\theta}$ values was set according to Viana (2011).

*2.3 Metamodeling framework for MC and SA runs*

Both MC simulations and SA are first conducted with the original models to obtain reference output data and reference computational times. Note that our objective is neither to perform an exhaustive uncertainty analysis using MC simulation nor to delve into the details of SA. Our focus here is to assess the metamodel-assisted analysis given that the composition model is computationally expensive to run these tasks. In particular, the MC simulation is used to investigate two things. First, the impact of using different structures of the Kriging metamodel on the approximation of original model output distribution. Second, the effect of the training sample size on the metamodel accuracy. For SA, a small number of parameters is involved in our case and thus we opt for the well-established method of Elementary Effects Test or Morris method (Morris, 1991). This approach is considered reasonable for the problem at hand while it allows for a manageable number of runs with the composition models in order to set a benchmark solution against the metamodel analysis.

An offline framework is employed where the metamodel is trained only once and then it is used for prediction. This is a generic framework which does not update the metamodel knowledge during the MC or the SA runs. It is convenient though for assessing the raw metamodel predictive skill without being affected by the selection of an infill strategy. A workflow of the present framework is presented below.

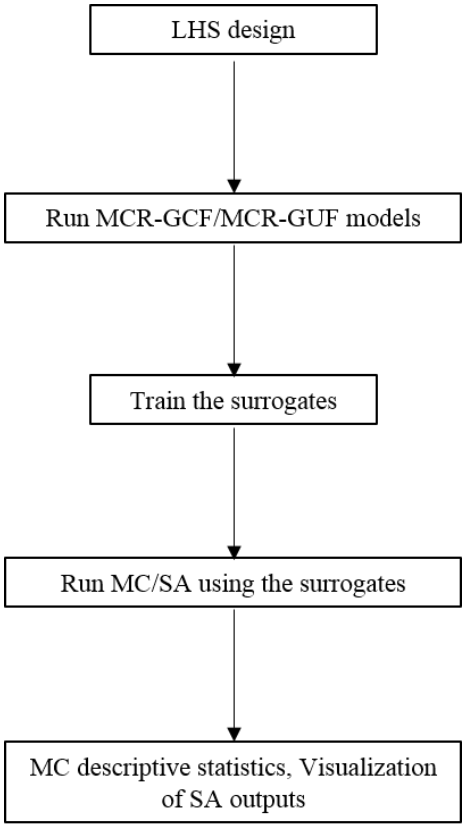


**Fig. 5.** Generic workflow of the offline metamodel framework.

The impact of the training sample size $n_t$ on the predictive power of an emulator is an important aspect of metamodeling and there are several suggestions in the literature related to the type of the metamodel (Razavi et al. 2012a). According to Jones et al. (1998) an empirical rule for the size of the training sample for Kriging metamodels could be in the order of $n_t = 10D$, with $D$ being the number of input variables. Obviously, this approach can lead to very large training samples as $D$ increases. In this study, the Kriging metamodels are initially trained based on a Latin hypercube sampling (LHS) (Mckay et al., 1979) with $10D$ design sites (in our case, $D = 3$). LHS was utilized due to its space-filling properties which allow for an effective exploration of the input parameter space while facilitating the smoothing interpolating capabilities of Kriging. Therefore, 30 training input patterns were generated using the LHS method and both the MCR-GCF and the MCR-GUF run on these to produce the required output. As an additional investigation on the impact of the training sample size, the same modelling procedure was repeated for 60 and 90 training points apart from the base case of 30 points.

The MC run is performed by randomly sampling the distributions of the input parameters. The hydraulic conductivity in both zones 1 and 2 of the groundwater model component, is considered as a lognormal random variable $Y = \ln(K)$ with mean value $\mu = 3.3241 m/day$ and variance $\sigma^2 = 0.1541$. In addition, recharge is assumed as an exponential random variable with mean value $1/\lambda$, where $\lambda = 1$. More specifically, a set of 1000 inputs was used in the case of the MCR-GCF model while a smaller set of 100 inputs was used in the case of the MCR-GUF model due to the computational restrictions discussed previously. For the MC simulation, the univariate scalar output $h$ represents the groundwater head on a specified observation point located at a region of the simulation field where significant variations are expected as the input parameters change. This point, denoted as O1, is located at the left corner of the simulation grid with coordinates$(X, Y) = (500, 300)$.

For the SA problem, we select an additional observation point, denoted as O2, located at $(X, Y) = (9400, 1100)$ where the hydraulic head differences are more pronounced between the MCR-GCF and the MCR-GUF models (Figure 6). In that sense, SA based on O1 enables the identification of the important parameters for the most "active" area of the simulation field. On the other hand, SA based on point O2 is suitable for associating the flow type differences with the input parameters. All model runs, pre- and post- processing tasks for both the original composition and the metamodels were implemented in MATLAB.

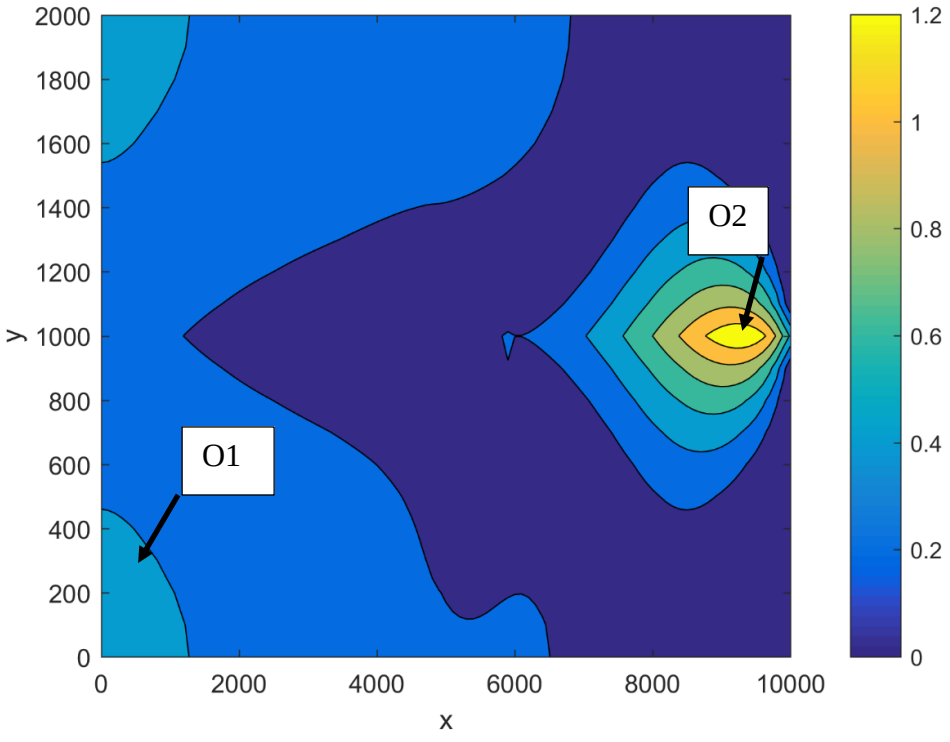


**Fig. 6.** Example of the absolute difference map between the simulation fields of MCR-GCF and MCR-GUF models at steady-state conditions.

## 3. RESULTS

### *3.1 MC simulation*

The prediction accuracy of the metamodels was quantified based on the MC output from the original models. For real-world problems, if the original model is so cheap to run that we can afford for MC simulations in the order of thousands the use of a metamodel would be redundant. Although the simulation time of the integrated models developed here could be reduced by setting appropriate initial conditions, our intention is to mimic the intensive runtimes of high-fidelity computer models for real-world applications. Since our study is of exploratory nature we attempt to assess the metamodels' accuracy using a large, yet, reasonable number of original model evaluations. As it can be seen in Table 1 even for the less computationally expensive MRC-GCF model, almost 3 days were required for a MC simulation of 1000 runs while the corresponding time with the Kriging metamodel was only 2.3 hr (including the original simulations for training the metamodel). Note that the MCR-GUF model is computationally expensive (single runtimes of ~ 63 minutes) and thus a smaller MC set of 100 runs was used to evaluate metamodels' performance.

**Table 1**

Computational times of the test simulation runs for both the original model and the metamodels.

| | **MCR-GCF** (1000 runs) | **Kriging (MCR-GCF)** (1000 runs) | **MCR-GUF** (100 runs) | **Kriging (MCR-GUF)** (100 runs) |
|---|---|---|---|---|
| **Computational time** (hr) | ~70 | ~2.3 | ~105.4 | ~31.5 |

The root mean square error (*RMSE*) metric was used to assess the differences between the different metamodel structures and the original models:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{ns}\left(h^{(i)} - \hat{h}^{(i)}\right)^2}{ns}} \quad (2)$$

where $\hat{h}$ denotes the metamodel prediction and $ns$ is the sample size of the validation set. In our case, $ns$ is the size of the input sample for the MC simulation. The *RMSE* values for the metamodel-based MC simulation for both composition cases are summarized in Table 2. The results refer to different combinations of regression and correlation functions of the metamodel structure for the base case scenario of 30 training patterns.

**Table 2**

*RMSE* values for metamodel prediction accuracy for $n_t = 30$ (lowest is highlighted).

| | MCR-GCF (1000 runs) Correlation model | | | MCR-GUF (100 runs) Correlation model | | |
|---|---|---|---|---|---|---|
| Polynomial order | Exponential | Gaussian | Linear | Exponential | Gaussian | Linear |
| 0 | 0.9702 | **0.5049** | 1.0355 | 0.8577 | **0.5924** | 1.7388 |
| 1 | 0.6559 | 0.5275 | 0.9676 | 0.7294 | 0.9474 | 0.8824 |
| 2 | 0.8393 | 0.7425 | 0.8001 | 0.6766 | 0.6808 | 0.7754 |

The low *RMSE* values indicate a good agreement between the MC test data and the metamodel predictions. The differences among the metamodel structures are relatively small, yet, the best results are obtained using a Gaussian-zero-order Kriging structure for both the MCR-GCF model and the MCR-GUF model. The Gaussian correlation function is commonly selected in most Kriging metamodeling studies, however, there are no systematic guidelines for selecting the best correlation model and results may be problem-dependent (Simpson and Martin, 2004; Song et al., 2013). In addition, we explore if the best metamodel structure changes with increasing the training points. Thus, the *RMSE* values were recalculated for each metamodel structure for different training samples and against the same MC output from the original models. In the case of the MCR-GCF model, an Exponential-2nd-order and an Exponential-1st-order were identified as best Kriging metamodel structures for the 60 and 90 training points, respectively. Interestingly, the Gaussian-zero-order structure was identified again as best to emulate the MCR-GUF model response and provided the lowest RMSE values for both the 60-point and the 90-point training samples. In general, the Exponential and Gaussian structures provided the best results, although the formulation of the Kriging

metamodel regarding the linear trend varied with the selection of different training samples in the case of the MCR-GCF model.

**Table 3**

*RMSE* values for metamodel prediction accuracy for $n_t = 60$ (lowest is highlighted).

| | MCR-GCF (1000 runs) Correlation model | | | MCR-GUF (100 runs) Correlation model | | |
|---|---|---|---|---|---|---|
| | Exponential | Gaussian | Linear | Exponential | Gaussian | Linear |
| Polynomial order | | | | | | |
| 0 | 1.0989 | 0.7435 | 1.0508 | 1.1763 | **0.5598** | 1.0496 |
| 1 | 0.6476 | 0.7857 | 0.7385 | 0.8519 | 0.6530 | 1.0394 |
| 2 | **0.4228** | 0.5413 | 0.5777 | 0.5983 | 0.7717 | 0.5738 |

**Table 4**

*RMSE* values for metamodel prediction accuracy for $n_t = 90$ (lowest is highlighted).

| | MCR-GCF (1000 runs) Correlation model | | | MCR-GUF (100 runs) Correlation model | | |
|---|---|---|---|---|---|---|
| | Exponential | Gaussian | Linear | Exponential | Gaussian | Linear |
| Polynomial order | | | | | | |
| 0 | 0.4859 | 0.4601 | 0.4437 | 0.3479 | **0.2891** | 0.3252 |
| 1 | **0.4175** | 2.9454 | 0.4991 | 0.3844 | 2.2393 | 0.3841 |
| 2 | 0.5073 | 2.6800 | 0.6246 | 0.5139 | 1.8571 | 0.6373 |

The performance of the best metamodel structures against the original model groundwater head data from the MC simulation is also visually demonstrated in Figure 7. It is evident that as more points are added for training, the metamodels' prediction converges to the original MC data.

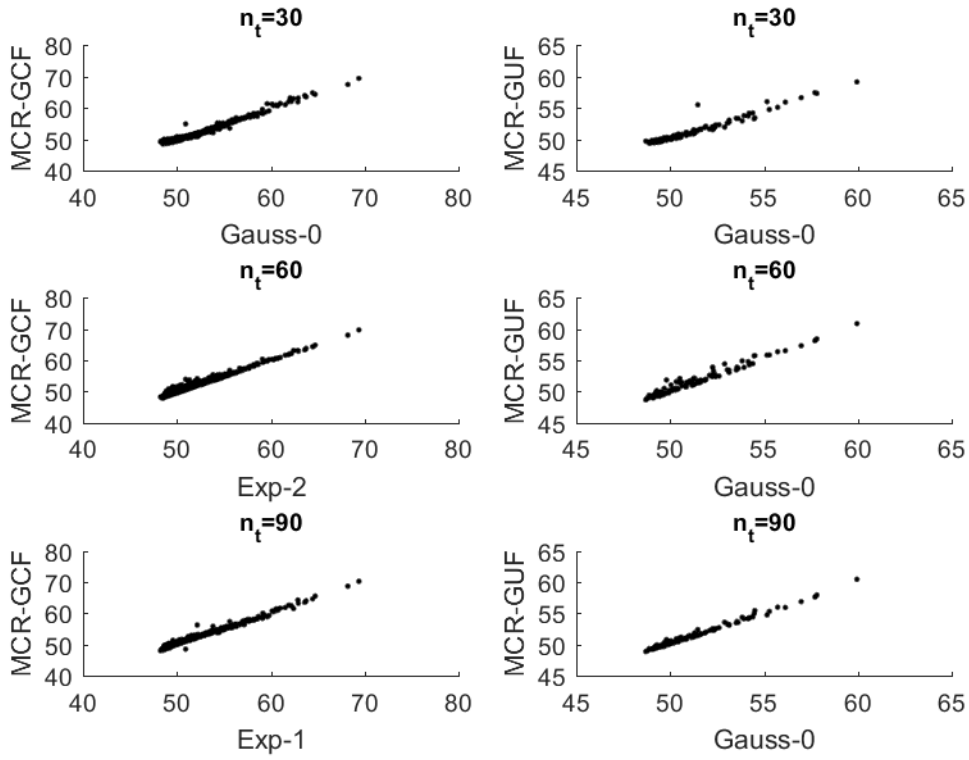


**Fig. 7.** Metamodel predictions versus the original model MC output for different training sample sizes. Left column refers to MCR-GCF model and right column to MCR-GUF model.

Frequency histograms were also utilized to provide information about the hydraulic head distribution from the original model and the metamodels' predictions (Figure 8). Again, the best metamodels are shown for the three different sets of training samples. However, a more conclusive result about the metamodels' performance can be obtained for the case of the MRC-GCF model where the MC set comprises 1000 runs. In general, the metamodels appear to reproduce adequately the original model distribution of the higher groundwater head values. In both cases though, the variability of the original model predictions is less well explained by the metamodels that were trained with 30 points (top row of figure 8). These metamodels mostly failed to capture the variability in the lower groundwater head values. From the histogram comparison it can be also implied that the metamodels are less successful to capture the MRC-GUF model behaviour which involves the non-linear unconfined aquifer flow. For example, the metamodel with a 60-point training sample, which is already 60% of the MC simulation budget for the MRC-GCF model, provides a poorer representation of the original hydraulic head distribution than the corresponding metamodel for the MRC-GCF model. However, the metamodels trained with larger data sets (60 and 90 points) appear to better reproduce the

distribution of the simulated heads from the original models. Note that MRC-GUF model only involves a set of 100 runs for the MC simulation due to the prohibitive computational cost.

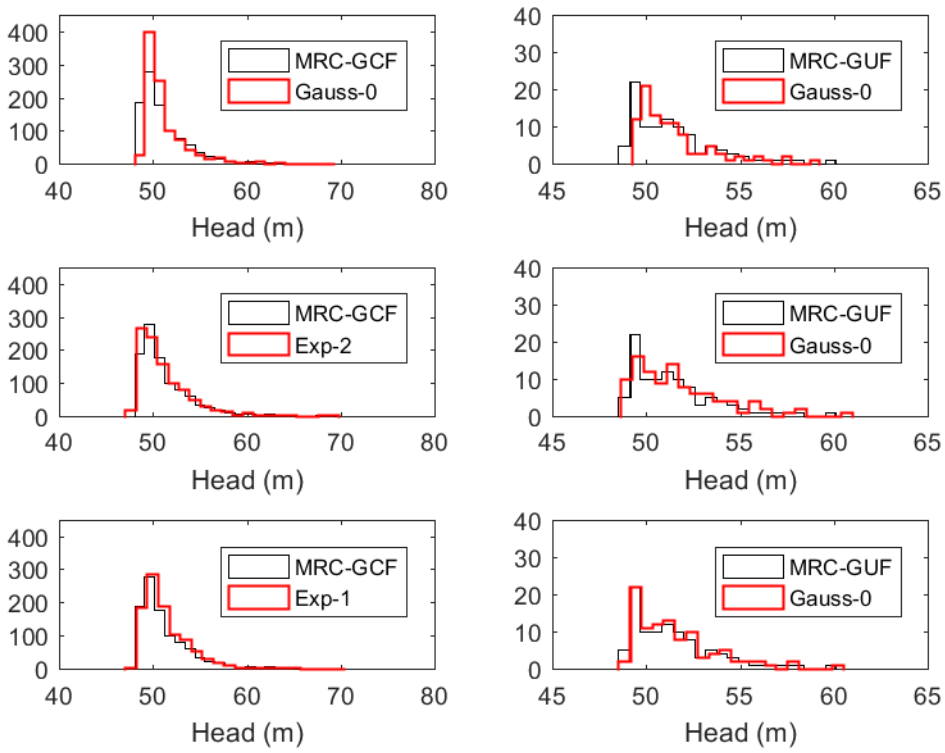


**Fig. 8.** Histograms of the MC output from the original models and the corresponding best metamodels for training sample sizes of 30, 60 and 90 training points.

*3.2 SA results*

The computational budget specified for SA studies, is related to both the type and the sampling approach of the SA methodology. For example, variance-based SA methods require a large number of model evaluations in the order of $1000 \times D$ (Pianosi et al., 2016). In our case $D$ is small, yet, a robust analysis would require 3000 model runs which leads to the unmanageable computational time of 125 days for the case of the computationally expensive MCR-GUF model. User's experience with the problem at hand could inform the SA approach by narrowing the parameter bounds or by eliminating inactive parameters. However, most high-fidelity, physics-based computer models involve a moderate to large number of parameters and/or external forcing variables with varying space-time behaviour so that a global SA would normally lead to thousands of model runs. In such cases, the use of a metamodel

may allow for more complex SA methods to be applied and the analysis of high-dimensional input vectors.

Here, due to the small number of input factors ($K_1$, $K_2$, $R_g$) and the long runtimes of the composition, we opt for the established method of Morris or Elementary Effect Test (EET), in order to screen important factors and identify possible interactions (Morris, 1991). EET may provide a reasonable SA output in the order of 10 to $100 \times D$ model runs (Pianosi et al., 2016). Since our aim is to compare the effectiveness and efficiency of the metamodels against the original model, EET, is a manageable SA framework in terms of computational requirements, particularly for the expensive MRC-GUF model.

The robust and well-documented SAFE Toolbox, developed by Pianosi et al. (2015), was used for the SA task. The EET for the original models was performed using $L = 8$ and $r = 10$ for the number of levels in the grid size and number of Elementary Effects, respectively (Pianosi et al., 2015). The total number of model runs with the Morris method is defined as $N_r = r \times (D + 1)$, thus, in our case $N_r = 40$. This amounts to a computational cost of 42 hr with the MRC-GUF model and 2.8 hr with the MRC-GCF model. The metamodel-based EET run, was based on the metamodels that were trained with $n_t = 30$ in order to approximately match the number of SA runs $N_r$ performed with the original model compositions. Given the negligible metamodel's evaluation time compared to the original models, we set $r = 75$ since we can easily afford for $N_r = 300$. In addition, we also assume that the same Kriging model structure is the best metamodel for the observational point O2 and it is fitted to 30 training points obtained for that point.

Figures 9 and 10 present the SA results related to the observation point O1 and O2, respectively, for the compositions and the corresponding metamodels. Qualitatively, all models identified similarly the important input factors and the possible interactions among the parameters. $R_g$ is the variable with the largest sample mean in all plots having the most significant impact on the simulated groundwater levels. The SA results are also consistent with the expected hydraulic behaviour of the models. In particular, $K2$ obviously has a stronger impact than $K1$ on the response of both MRC-GCF and MRC-GUF models at O1, since the latter is located in the part of the simulation grid where $K2$ is assigned (see Figure 1). Perhaps unsurprisingly, the opposite effect of $K2$ is observed for the observation point O2.

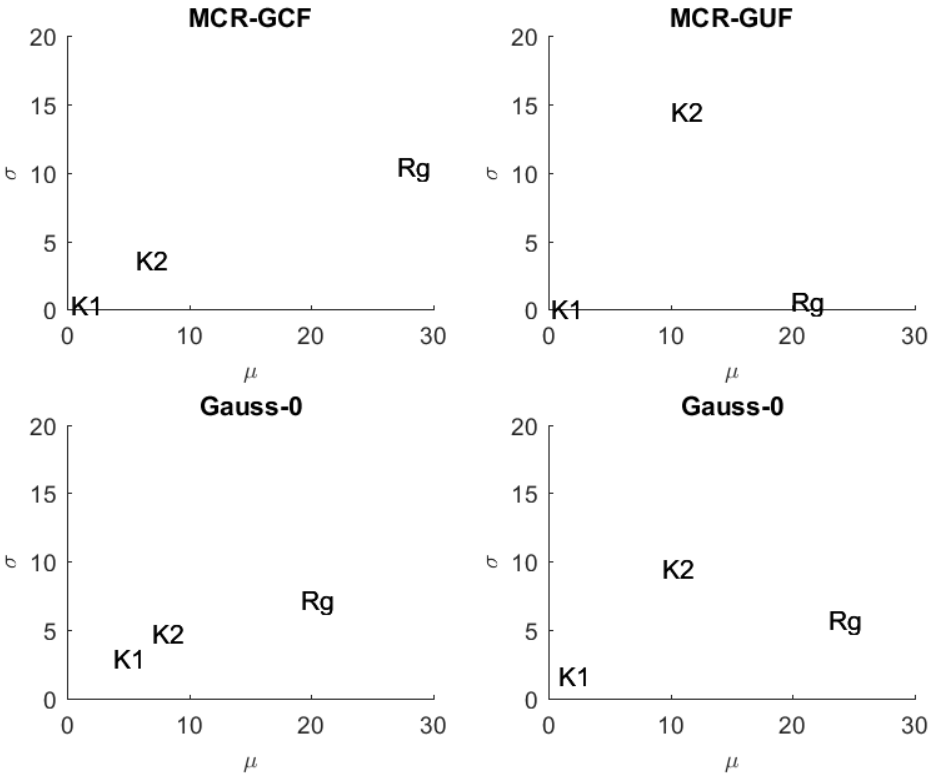


**Fig. 9.** Sample means $\mu$ versus sample standard deviations $\sigma$ of EET. The results refer to the observation point O1.

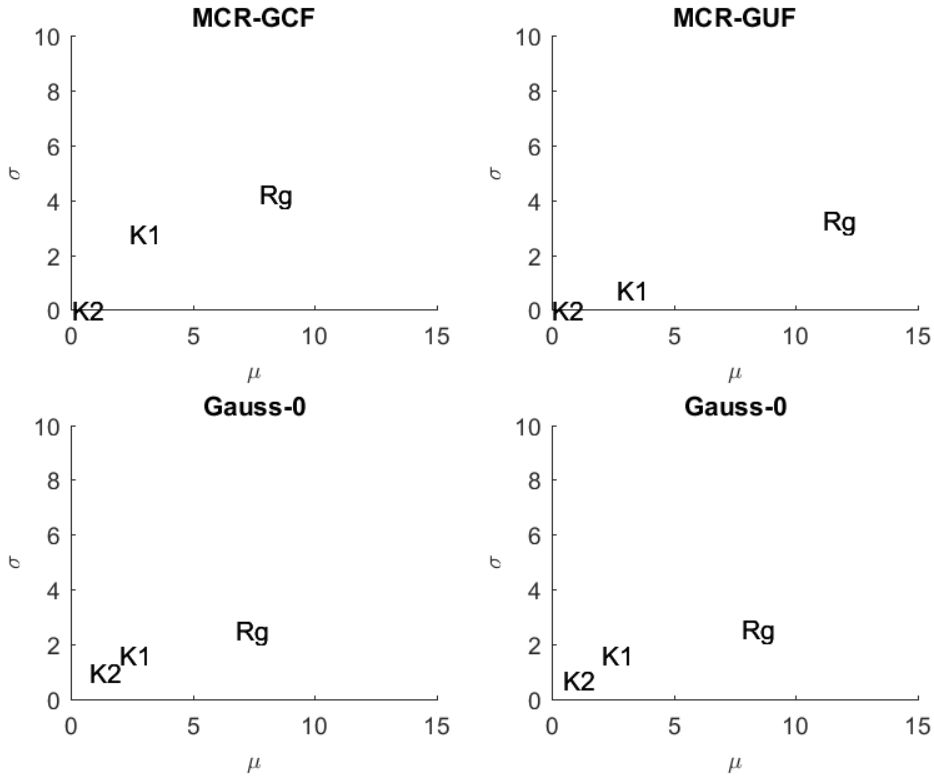


**Fig. 10.** Sample means $\mu$ versus sample standard deviations $\sigma$ of EET. The results refer to the observation point O2.

However, there are some discrepancies between the compositions and the metamodels. In the case of the MCR-GCF model at observation point O1, $R_g$ appears to have a stronger impact and a larger degree of interaction with other parameters compared to the metamodel results. For the same observation point, EET based on the MCR-GUF model shows that $K2$ is involved in interactions whereas $R_g$ is not. The metamodel-based EET gives a low but yet measurable degree of interaction for $R_g$. On the other hand, at observation point O2, the metamodel-based EET is more similar to the composition analysis. $K1$ and $K2$ are grouped together and more clustered around the origin as shown by both compositions and metamodels. This difference can be explained by the observation point O2 being hydraulically more controlled than O1 due to the river impact and thus less variability of hydraulic heads is expected. Therefore, the metamodels were able to capture the response of the compositions to different inputs better than O1, for $n_t = 30$.

Since O1 is expected to be more sensitive to the input variables a second SA was performed for the MCR-GCF model by increasing $r$ to 75 and thus resulting in a number of 300 runs ($100 \times D$) for the EET. That was to identify if both the composition and the emulator misplaced the importance or the interaction of the input variables due to the low amount of original model runs. That is, 40 runs for SA with MCR-GCF model and only 30 training points to train the Kriging metamodel. The SA results from the MCR-GCF model are compared with the best metamodel found using 90 training points, which are presented in Table 4. Note that in real-world problems 300 runs may be computationally impractical, for example, it would take approximately 13.5 days to run the MRC-GUF model. In this case, the SA based on the MCR-GCF model required a computational time of 21.2 hr while the metamodel-based SA, including the training time, required 6.3 hr. The results from the SA based on 300 runs with the MRC-GCF model are presented in Figure 11.

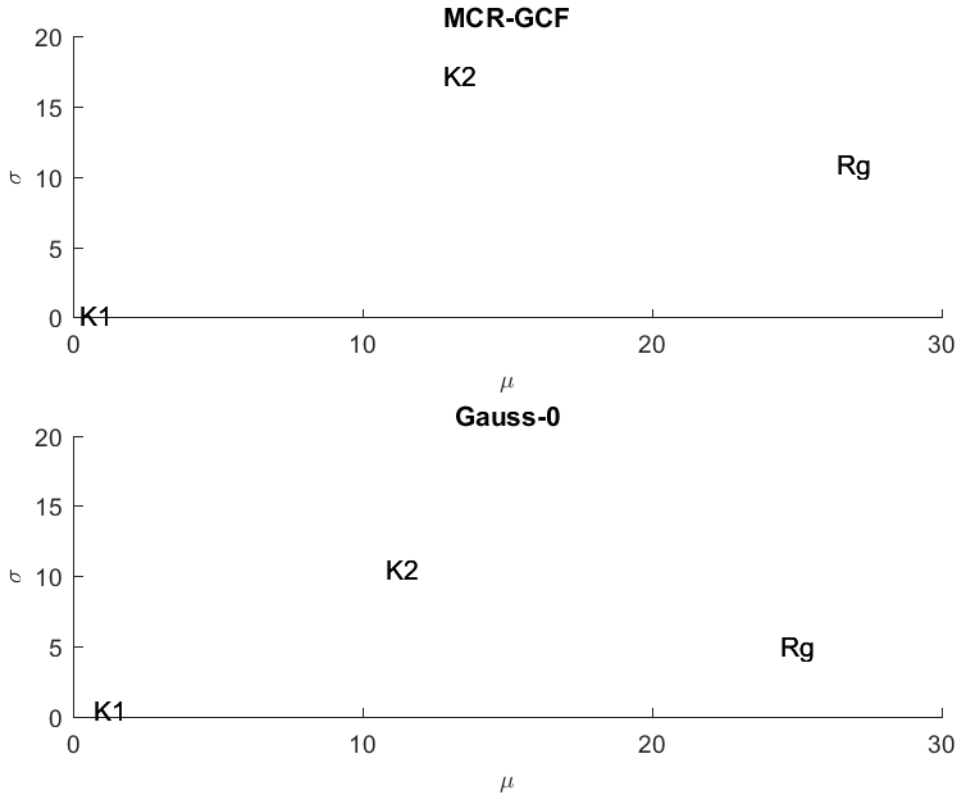


**Fig. 11.** Sample means $\mu$ versus sample standard deviations $\sigma$ of EET, produced by the MCR-GCF model after increasing the number of runs to 300. The corresponding metamodel results are also shown.

Results demonstrate a different placement of the input variables with the increase in the number of simulations. The sample means and the sample standard deviations of $K1$ and $R_g$ are very similar between the metamodel and the composition. $K2$ is identified as the variable mostly involved in interactions with the metamodel-based EET showing a smaller measure of spread. In overall, the comparison between the original model and the metamodel indicates a good match in terms of screening the important parameters while reducing the computational budget by 70%.

## 4. CONCLUSIONS

Integrated modelling is a promising, relatively new approach for solving complex environmental problems. It is based on the concept of linking together stand-alone codes that simulate different aspects of the system under study. One possible challenge associated with integrated modelling is the resulting computational burden when either complex models are

linked together and/or dynamic processes are simulated. In that case, computationally demanding tasks which require multiple runs of the integrated model may be hampered by impractical runtimes. To address that, the use of metamodels was proposed in the present study in order to perform MC simulations and SA in manageable computational times.

In particular, we employed Kriging metamodels for emulating the steady-state response of a distributed groundwater model that was linked together with a simple river model via the OpenMI software platform. Two compositions were implemented, one with a confined groundwater flow model that reached steady-state after 4.2 minutes and one with an unconfined groundwater flow model that reached steady-state after 63 minutes. First, the compositions run to get benchmark solutions for MC simulations and SA. Then, the metamodels were trained using a small number of input-output data based on Latin Hypercube Sampling and used in lieu of the composition models to perform the same tasks.

Results from the MC simulations demonstrate that the predictions of the metamodels were in a good agreement with the responses of the original models while the overall computational savings were in the order of 90% and 70% in the case of the confined groundwater model (MCR-GCF) and the unconfined (MCR-GUF), respectively. In our study, the impact of the metamodel structure was not decisive regarding its prediction skills. However, there were specific combinations of covariance and regression models that provided the best results. In general, in this study, Gaussian and Exponential correlation models provided the lowest error values against the original models. While an increase of the training sample size, improved, as expected, the Kriging metamodel prediction skills, there was not a specific Kriging structure that was consistently identified as the best for MCR-GCF model. However, a Gaussian correlation model was identified as best for the MCR-GUF model. It should also be noted that the Kriging metamodels provided a poorer representation of the original output distribution in the case of the composition with the unconfined groundwater flow model. The unconfined groundwater flow model is characterized by a stronger non-linear behaviour than the confined model which is not in favour of constructing robust metamodels of global accuracy, particularly when the metamodel is used in an offline training framework.

For the SA runs, it was observed that by using only a small amount of training data, the metamodels comparably screened the input parameters and those that can interact with others with the SA based on the compositions. The results were consistent with the expected hydraulic response of the groundwater models given the specified parameter input set. Due to their fast implementation the metamodel-based SA was performed easily in the order of the recommended number of runs for an EET approach without worrying about the computational

cost A more conclusive result was obtained for the MCR-GUF model where its runtimes allowed for an EET at the order of $100 \times D$. By comparing the results produced by the metamodel and the original model for the less hydraulically controlled observation point, the input variables were screened in a similar way while the overall computational time was reduced by 70%.Overall, Kriging metamodels provided a fast and reasonably accurate approximation to the computationally expensive integrated models. Whilst, the present study represents a simple example of the benefits of metamodeling to integrated modelling, based on the promising results presented here, future work is oriented to the emulation of much more complex compositions.

Reducing the computation time that integrated modelling (IM) takes is one of the key challenges in making IM usable to address deceptively simple questions. There are a number of ways of doing this: HPC "brute force and ignorance", parallelisation and emulation. There are advantages and disadvantage for all three approaches. HPC requires learning a new language / porting an existing composition to another language, e.g. UNIX; parallelisation enables different components from the composition to be run on different processors or batches of runs for multiple realisations for MC techniques. Emulation, as shown in this paper, requires investment in time to develop the reduced complexity models, but reaps rewards in terms of reduced runtime, particularly for multiple realisations.

By representing highly computational demanding linked model systems by emulators then sensitivity and uncertainty analysis can be undertaken. The next steps are using and testing the techniques on more complex compositions: those containing more model instances and greater complexity of feedbacks between the different components. Given that model compositions can be made up of a range of different components including databases, real-time feeds from sensors and models then emulators can themselves be components in compositions. This adds to the flexibility of any model linking approach and ensures that integration can be achieved even if the outputs are from complex, computationally demanding model instances.

**Acknowledgements**

We thank the commissioning editor and the two anonymous reviewers for their helpful comments in improving the paper. The authors publish with the permission of the Executive Director of the British Geological Survey (NERC).

## References


Aly A.H., Peralta R.C., 1999. Optimal design of aquifer cleanup systems under uncertainty using a neural network and a genetic algorithm. Water Resour. Res., 35(8), 2523-2532.

Asher M. J., Croke B.F.W., Jakeman A.J., Peeters L.J.M., 2015. A review of surrogate models and their application to groundwater modeling. Water Resour. Res., 51(8), 5957-5973.

Babaei M., Pan I., Alkhatib A., 2015. Robust optimization of well location to enhance hysteretical trapping of $CO_2$: Assessment of various uncertainty quantification methods and utilization of mixed response surface surrogates. Water Resour. Res., 51(12), 9402-9424.

Baú D.A., Mayer A.S., 2006. Stochastic management of pump-and treat-strategies using surrogate functions. Adv. Water Resour. 29, 1901-917.

Barkwith A.K.A.P., Pachocka M., Watson C., Hughes A.G., 2014. Couplers for linking environmental models: scoping study and potential next steps. Nottingham, UK, British Geological Survey. (OR/14/022) (Unpublished)

Bastin L., Cornford, D., Jones R., Heuvelink G.B., Pebesma E., Stasch C., Nativi S., Mazzetti P., Williams M., 2013. Managing uncertainty in integrated environmental modelling: The UncertWeb framework. Environ. Modell. Softw., 39, 116-134.

Bellos V., Kourtis I.M., Moreno-Rodenas A., Tsihrintzis V.A., 2017. Quantifying roughness coefficient uncertainty in urban flooding simulations through a simplified methodology. Water, doi:10.3390/w9120944.

Bianchi M., Zheng L., Birkholzer J.T., 2016. Combining multiple lower-fidelity models for emulating complex model responses for CCS environmental risk assessment. Int. J. Greenh. Gas Con, 46, 248-258.

Broad D.R., Dandy G.C., Maier H.R., 2014. A systematic approach to determining metamodel scope for risk-based optimization and its application to water distribution system design. Environ. Modell. Softw., 69, 382-395.

Borgonovo E., Castaings W., Tarantola S., 2012. Model emulation and moment-independent sensitivity analysis: An application to environmental modelling. Environ. Modell. Softw., 34, 105-115.

Campolongo F., Cariboni J., Saltelli A., 2007. An effective screening design for sensitivity analysis of large models. Environ. Modell. Softw., 22,1509-1518.

Castelletti A., Pianosi F., Soncini-Sessa R., Antenucci J.P., 2010. A multiobjective response surface approach for improved water quality planning in lakes and reservoirs. Water Resour. Res., doi:10.1029/2009WR008389.

Castronova A.M., Goodall J.L., Elag M.M., 2013. Models as web services using the open geospatial consortium (ogc) web processing service (wps) standard. Environ. Modell. Softw., 41, 72-83.

Christelis V., Mantoglou A., 2016. Pumping optimization of coastal aquifers assisted by adaptive metamodeling methods and radial basis functions. Water Resour. Manag., 30(15), 5845-5899.

Christelis V., Bellos V., Tsakiris G., 2016. Employing surrogate modelling for the calibration of a 2D flood simulation model. In Sustainable Hydraulics in the Era of Global Change, Proceedings of 4th IAHR Congress, Liege, Belgium, 27–29 July 2016; Erpicum, S., Dewals, B., Archambeau, P., Pirotton, M., Eds.; CRC Press: Boca Raton, FL, USA, 2016; pp. 727–732.

Christelis V., Mantoglou A., 2017. Physics-based and data-driven surrogate models for pumping optimization of coastal aquifers. European Water 57:481-488.

Christelis V., Regis R.G., Mantoglou A., 2018. Surrogate-based pumping optimization of coastal aquifers under limited computational budgets. J. Hydroinform., doi: 20(1), 164-176.

Coutts S.R., Yokomizo H., 2014. Meta-models as a straightforward approach to the sensitivity analysis of complex models. Popul. Ecol., 56(1), 7-19.

Fienen M.N., Nolan B.T., Feinstein D.T., 2016. Evaluating the sources of water to wells: Three techniques for metamodeling of a groundwater flow model. Environ. Modell. Softw., 77, 95-107.

Forrester A.I.J., Sóbester A., Keane A.J., 2008. Engineering design via surrogate modelling-A practical guide. Wiley, New York.

Forrester A.I.J., Keane A.J., 2009. Recent advances in surrogate-based optimization. Prog. Aerosp. Sci., 49, 50-79.

Ge Q., Ciuffo B., Menendez M., 2015. Combining screening and metamodel-based methods: An efficient sequential approach for the sensitivity analysis of model outputs. Reliab. Eng. Syst Safe, 134, 334-344.

Gregersen J.B., Gijsbers P.J.A., Westen S.J.P., 2007. OpenMI: Open modelling interface. J. Hydroinform., 9(3), 175-191.

Harpham Q., Cleverley P., Kelly D., 2014. The FluidEarth 2 implementation of OpenMI 2.0. J. Hydroinform., 16(4), 890-906.

Harpham Q., Danovaro E., 2015. Towards standard metadata to support models and interfaces in a hydro-meteorological model chain. J. Hydroinform., 17(2), 260-274.

Harpham Q., Lhomme J., Parodi A., Fiori E., Jagers B., Galizia A., 2016. Using OpenMI and a Model MAP to Integrate WaterML2 and NetCDF Data Sources into Flood Modeling of Genoa, Italy. JAWRA Journal of the American Water Resources Association, 52(4), 933-949.

Jackson C.R., Spink A.E.F., 2004. User's manual for the groundwater flow model ZOOMQ3D.

Jones D., Schonlau M., Welch W., 1998. Efficient global optimization of expensive black-box functions. J. Glob. Optim. 13, 455–492.

Kleijen J.P.C., 2009. Kriging metamodeling in simulation: A review. Eur. J. Oper. Res., 192, 707-716.

Krige DG (1951) A statistical approach to some basic mine valuation problems on the Witwatersrand. J. Chem. Metall. Min. Eng. Soc. S. Afr. 52(6), 119–139.

Khu S.T.K., Werner M.G.F., 2003. Reduction of Monte-Carlo simulation runs for uncertainty estimation in hydrological modelling. Hydrol. Earth Syst. Sc., 7(5), 680-692.

Laniak G.F., Olchin G., Goodall J., Voinov A., Hill M., Glynn P., Whelan G., Geller G., Quinn N., Blind M., Peckham S., 2013. Integrated environmental modeling: a vision and roadmap for the future. Environ. Modell. Softw., 39, 3-23.

Lophaven S., Nielsen H., Sondergaard J., 2002. DACE - A Matlab Kriging Toolbox, Version 2.0. Technical Report IMM-TR-2002-12. Informatics and Mathematical Modelling, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark.

Makropoulos C., Safiolea E., Baki S., Douka E., Stamou A., Mimikou M., 2010. An Integrated, Multi-modelling Approach for the Assessment of Water Quality: Lessons from the Pinios River Case in Greece. International Environmental Modelling and Software Society (iEMSs). 2010 International Congress on Environmental Modelling and Software Modelling for Environment's Sake, Fifth Biennial Meeting, Ottawa, Canada.

Mansour M., Mackay J., Abesser C., Williams A., Wang L., Bricker S., Jackson C., 2013. Integrated Environmental Modeling applied at the basin scale: linking different types of models using the OpenMI standard to improve simulation of groundwater processes in the Thames Basin, UK.

McKay M.D., Conover W.J., Beckman R.J., A comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, Technometrics, 21(2), 1979.

Morris M., 1991. Factorial sampling plans for preliminary computational experiments.
Technometrics 33 (2), 161-174.

Nalbantis I., Efstratiadis A., Rozos E., Kopsiafti M., Koutsoyiannis D., 2011. Holistic versus monomeric strategies for hydrological modelling of human-modified hydrosystems. Hydrol. Earth Syst. Sc., 15, 743-758.

Nativi S., Mazzetti P., Geller G.N., 2013. Environmental model access and interoperability: The GEO Model Web initiative. Environ. Modell. Softw., 39, 214-228.

Peckham S.D., Hutton E.W., Norris B., 2013. A component-based approach to integrated modeling in the geosciences: The design of CSDMS. Comput. Geosci., 53, 3-12.

Pianosi F., Sarrazin F., Wagener T., 2015. A Matlb toolbox for Global Sensitivity Analysis. Environ. Modell. Softw., 70, 80-85.

Pianosi F., Beven K., Freer J., Hall J.W., Rougier J., Stephenson D.B., Wagener T., 2016. Sensitivity analysis of environmental models: A systematic review with practical workflow. Environ. Modell. Softw., 79, 214-232.

Queipo N.V., Haftka R.T., Shyy W., Goel T., Vaidyanathan R., Tucker P.K., 2005. Surrogate-based analysis and optimization. Prog. Aerosp. Sci., 41, 1-28.

Ratto M., Castelletti A., Pagano A., 2012. Emulation techniques for the reduction and sensitivity analysis of complex environmental models. Environ. Modell. Softw., 34, 1-4.

Rajabi M.M., Ataie-Astiani B., Simmons C.T., 2015. Polynomial chaos expansions for uncertainty propagation and moment independent sensitivity analysis of seawater intrusion simulations. J. Hydrol. 520, 101-122.

Razavi S., Tolson B.A., Burn D.H., 2012a. Review of surrogate modelling in water resources. Water Resour. Res., doi:10.1029/2011WR011527.

Razavi S., Tolson B.A., Burn, D.H., 2012b. Numerical assessment of metamodelling strategies in computationally intensive optimization. Environ. Modell. Softw., 34, 67-86.

Sacks J., Welch W.J., Mitchell T.J., Wynn H.P., 1989. Design and analysis of computer experiments. Stat. Sci., 4, 409-435.

Simpson T.W., Martin J.D., 2004. On the Use of Kriging Models to Approximate Deterministic Computer Models. In ASME 2004 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (pp. 481-492). American Society of Mechanical Engineers.

Shretha D.L, Kayastha N., Solomatine D.P., 2009. A novel approach to parameter uncertainty analysis of hydrological models using neural networks. Hydrol. Earth Syst. Sc., 13, 1235-1248.

Shoemaker, C. A., Regis, R. G., & Fleming, R. C. (2007). Watershed calibration using multistart local optimization and evolutionary optimization with radial basis function approximation. Hydrological sciences journal, 52(3), 450-465.

Song X., Zhang J., Zhan C., Xuan Y., Ye M., Xu C., 2015. Global sensitivity analysis in hydrological modelling: Review of concepts, methods theoretical frameworks and applications. J. Hydrol. 523, 739-757.

Song H., Choi K.K., Lamb D., 2013. A study on improving the accuracy of kriging models by using correlation model/mean structure selection and penalized log-likelihood function. 10th World Congress on Structural and Multidisciplinary Optimization. Florida, Orlando.

Storlie, C.B., Swiler, L.P., Helton, J.C., Sallaberry, C.J., 2009. Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. Reliab. Eng. Syst Safe, 94(11), 1735-1763.

Sun A.Y., Miranda R.M., Xu X., 2015. Development of multi-metamodels to support surface water quality management and decision making. Environ. Earth Sci., 73, 423-434.

Tsoukalas I., Kossieris P., Efstratiadis A., Makropoulos C., 2016. Surrogate-enhanced evolutionary annealing simplex algorithm for effective and efficient optimization of water resources problems on a budget. Environ. Modell. Softw., 77, 122-142.

Tsoukalas I., Makropoulos C, 2015a. Multiobjective optimisation on a budget: Exploring surrogate modelling for robust multi-reservoir rules generation under hydrological uncertainty. Environ. Modell. Softw., 69, 396-413.

Tsoukalas I., Makropoulos C, 2015b. A surrogate based optimization approach for the development of uncertainty-aware reservoir operational rules: the case of nestos hydrosystem. Water Resour. Manag., 29(13): 4719-4734.

Viana F.A.C., 2011. SURROGATES Toolbox, Gainesville, FL, USA, Ver. 3.0, available at https://sites.google.com/site/felipeacviana/surrogatestoolbox

Voinov A., Nagesh K., McCall K., Glynn P.D., Kragt M. E., Ostermann F.O., Pierce S. A., Ramu P., 2016. Modelling with stakeholders–next generation. Environ. Modell. Softw., 77: 196-220.

Wang C., Duan Q., Gong W., Ye A., Di Z., Miao C., 2014. An evaluation of adaptive surrogate modelling based optimization with two benchmark problems. Environ. Modell. Softw., 60, 167-179.

Yan S., Minsker B., 2006. Optimal groundwater remediation design using an adaptive neural network genetic algorithm. Water Resour. Res., doi:10.1029/2005WR004303.

Yazdi J., Salehi Neyshabouri S.A.A, 2015. Adaptive surrogate modelling for optimization of flood control detention dams. Environ. Modell. Softw., 61, 106-120.

Yu J.J., Qin X.S., Larsen O., 2015. Uncertainty analysis of flood inundation modelling using GLUE with surrogate models in stochastic sampling. Hydrol. Process. 29, 1267-1279.

Zou R., Lung W-S, Wu J., 2009. Multiple-pattern parameter identification and uncertainty analysis approach for water quality modelling. Ecol. Model., 220, 621-629.

Xiu D., Karniadakis G.E., 2002. The Wiener-Askey polynomials chaos for stochastic differential equations. SIAM, J. Sci. Comput., 24(2), 619-644.