Combining multiple lower-fidelity models for emulating complex model responses for CCS environmental risk assessment

Marco Bianchi[12], Liange Zheng[1*], Jens T. Birkholzer[1]

[1]Earth Sciences Division,
Lawrence Berkeley National Laboratory,
Berkeley, California, 94720, USA

[2] now at British Geological Survey,
Kingsley Dunham Centre,
Keyworth, Nottingham, NG12 5GG, UK

*Corresponding author:
Liange Zheng, Earth Sciences Division, Lawrence Berkeley National Laboratory (LBNL),
Berkeley, CA 94720, USA. (LZheng@lbl.gov)
Phone: 510-486-5502

Abstract. Numerical modeling is essential to support natural resource management and environmental policy-making. In the context of CO_2 geological sequestration, these models are indispensible parts of risk assessment tools. However, because of increasing complexity, modern numerical models require a great computational effort, which in some cases may be infeasible. An increasingly popular approach to overcome computational limitations is the use of surrogate models. This paper presents a new surrogate modeling approach to reduce the computational cost of running a complex, high-fidelity model. The approach is based on the simplification the high-fidelity model into computationally efficient, lower-fidelity models and on linking them with a mathematical function (linking function) that addresses the discrepancies between outputs from models with different levels of fidelity. The resulting linking function model, which can be developed with small computational effort, can be efficiently used in numerical applications where multiple runs of the original high-fidelity model are required, such as for uncertainty quantification or sensitivity analysis. The proposed approach was then applied to the development of a reduced order model for the prediction of groundwater quality impacts from CO_2 and brine leakage for the National Risk Assessment Partnership (NRAP) project.

Key Words: Risk assessment; Model; Reduced Order model; Groundwater; surrogate modeling

1 Introduction and background

Despite the outstanding and consistent progress in computational efficiency, a systematic application of computer simulations to support natural resource management and environmental policy-making is still limited because of the great computational effort required to run modern environmental models. Advances in computational power, together with progresses in scientific knowledge, have in fact pushed for the development of more and more complex models with an increasingly larger number of processes and input parameters. The toll for the improved realism of these modern complex models is paid in terms of execution time, which can easily become practically infeasible when the spatial and/or temporal scale of the natural system is large, or when a large number of model responses needs to be calculated. The latter is typical of important numerical applications such as automatic model calibration, multi-objective optimization, sensitivity analysis, and uncertainty quantification.

CO_2 geologic storage is being considered as a possible measure to curb the anthropogenic emissions of greenhouse gases. A careful assessment of the risks associated with CO_2 geologic storage is critical to deployment of large scale CO_2 geological storage. One of the potential risks is the impact of CO_2 leakage from deep subsurface reservoirs into overlying groundwater aquifers. Therefore, contamination of groundwater due to leakage in shallow aquifers is considered one of the major risks considered in risk profiles developed by the National Risk Assessment Partnership (NRAP) project, a program that quantifies the behavior of engineered-natural system for CO_2 storage and uses science-based predictions to inform decisions tied to CO_2 geological sequestration. Numerical models for evaluating the impact of CO_2 leakage on groundwater, a process involving multiphase flow and reactive transport with complex chemical reactions, are very complex and also involve large uncertainties. Therefore, more computationally efficient models are needed for the development of risk profiles.

An increasingly popular approach to overcome computational limitations is the use of surrogate models (i.e., reduced order models, metamodels, emulators, and lower-fidelity models), which represent simplified and faster-to-run models that mimic (emulate) the output of the original model for a specified set of input parameters. Surrogate modeling has been applied in several scientific and engineering disciplines mainly in support of engineering design optimization and calibration (*Simpson et al.,* 2001, 2008, *Jones*, 2001, *Queipo et al.*, 2005, *Wang and Shan*, 2007, *Forrester and Keane* 2009, *Forrester,* 2010). In the context of environmental sciences, surrogate models have been used used for performing sensitivity analysis and calibration of complex models (*Liong et al*., 2001; *Mugunthan et al*. 2005; *Bliznyuk et al.*, 2008; *Matott and Rabideau*, 2008; *Ratto et al*., 2012; *Sun et al*., 2012), design of groundwater wells and pumping management (*Hemker et al*., 2008; *Kourakos and Mantoglou*, 2009; *Sreekanth and Datta*, 2011), and optimization of groundwater and soil remediation systems (*Baú and Mayer*, 2006; *Regis and Shoemaker*, 2007, 2009; *Fen et al.*, 2009). In NRAP, reduced order models (ROMs) simulating CO_2 transport in reservoir, wellbore leakage (e.g. Jordan et al., 2015) and groundwater contamination (Dai et al., 2014) were included in a system tool for estimating the long-term risks of CO_2 sequestration projects (Pawar, et al., 2014) .

Surrogate models can be classified into two broad categories [*Razavi et al*., 2012a]. The first is represented by response surface models based on data-driven functions that empirically emulate the output of the original model (e.g., *Dyn et al*., 1986; *Sacks et al*., 1989; *McKay*, 1991; *Myers and Montgomery*, 1995). These functions are developed by fitting a set of original model runs at specific points or design sites in the input parameter space. For instance, the ROMs developed in NRAP belong to this category. The second category of surrogate models includes simplified, physically-based lower-fidelity models that are used in place of a computationally demanding model. In this context, the original model is usually designated as the "high-fidelity" model and the term "fidelity" is intended as the ability to represent the system of interest. In comparison with response surface surrogates, lower-fidelity surrogates provide more accurate results in those regions of the input parameter space that do not include a large number of design sites (*Razavi et al*., 2012a). Moreover, this type of surrogate models are not afflicted by the problem of dimensionality, which limits the application of response surface surrogates to problems with a large number of input parameters (e.g., *Koch et al.*, 1999, *Simpson et al,*. 2008).

Lower-fidelity surrogate modeling has been mostly applied to reduce the computational load of optimization problems (*Alexandrov et al.*, 2001; *Vitali et al.*, 2002; *Eldred et al*., 2004; *Gano et al*., 2006; *Robinson et al*., 2006; *Forrester et al.*, 2007; *Forrester and Keane*, 2009; *Sun et al.*, 2010; *Berci et al*., 2011; *Koziel and Leifsson*, 2012). With this approach, known as "multi-fidelity" or "variable-fidelity" optimization in the literature, the difference or the ratio between outputs from high-fidelity and lower-fidelity models is simulated with a correction function usually represented by a polynomial (*Madsen and Langthjem*, 2001; *Viana et al.*, 2009; *Sun et al*., 2010), but also modeled with other approaches such as kriging (*Huang et al.*, 2006; *Gano et al*., 2006; *Forrester et al.*, 2007; *Kleijnen*, 2009) and neural networks (*Leary et al*., 2003; *Kim and Koc*, 2007; *Sun et al*., 2010).

Despite the considerable amount of literature about surrogate models and the development of several different approaches for coupling lower-fidelity and high-fidelity models, very few studies have considered systems that can be simulated with more than two levels of fidelity. Correction functions used in multi-fidelity optimization are in fact typically designed to model the discrepancies between the high-fidelity model and a single lower-fidelity model. The only exceptions include the co-kriging approach presented by *Forrester et al.* (2007), and the Bayesian Gaussian process model introduced by *Kennedy and O'Hagan* (2000), and subsequently extended by *Quian et al*. (2006, 2008) and by *Goh et al*. (2012). However, these approaches can be extended to multiple lower-fidelity models by assuming a hierarchical combination of models. In other words, outputs from the model with the highest fidelity can be written as a combination of the functions describing the discrepancies between each pair of lower-fidelity models. Moreover, the majority of surrogate modeling approaches is applicable when the number of input parameters (known as the input parameter space)in the high-fidelity model is the same as in the lower-fidelity model. The development of methods for handling multi-fidelity models with different input parameter spaces has received very little attention in the surrogate modeling literature (*Simpson et al*., 2008).

In this work we propose a surrogate modeling approach to emulate the output of a high-fidelity model from the outputs of a number of independent (i.e., not hierarchical), lower-fidelity models or ROMs. This approach, which we refer to as Linking Function Surrogate Modeling (LFSM), can be particularly effective to emulate the response of large-scale environmental models considering several physical and chemical processes. These models are commonly used in risk assessment in Carbon Capture and Storage applications. It is in fact common, especially in the simulation of natural systems, that each of these processes can be simulated separately (provided these processes are not tightly coupled), even if the highest level of realism is achieved with a high-fidelity model that couples all the relevant processes in its mathematical formulation. Unfortunately, the use of such a high-fidelity model in computationally expensive numerical analysis (e.g., global sensitivity analysis, uncertainty quantification, optimization, and risk assessment) is often computationally infeasible. The proposed approach can be a valid support to significantly reduce execution time without compromising the realism and the accuracy of the simulation. Even though our focus is on simulations of multiphase flow and solute transport in geological media, the method is general and it can be applied in other scientific and engineering disciplines.

1.1 Role of surrogate modelling in NRAP

NRAP is adapting and building system platforms for performing integrated assessment modeling of CO_2 storage sites. Surrogate models (i.e., ROMs) that provide reliable results in a small fraction of the time required to run complex process-based numerical simulations are required to assess the risk of CO_2 leakage in shallow groundwater. To overcome the difficulty of deriving such surrogate models from multiple runs of high-fidelity numerical models considering 3-D heterogeneous multiphase flow and reactive transport processes, two separate ROMs are used to represent the complex hydrogeological and geochemical conditions in a heterogeneous aquifer. The first ROM was developed from a numerical model that accounts for the heterogeneous flow and transport conditions in the presence of multiple leakage wells. The second ROM was obtained from numerical models that feature greatly simplified flow and transport conditions, but allow for a more complex representation of all relevant geochemical reactions. Clearly, neither ROM can separately provide an accurate prediction of the risk profile, because of the simplifications inherent in these models. The proposed LFSM provides an alternative approach that allows linking the outputs from two separate ROMs to calculate reliable predictions of the volume of aquifer impacted by leakage of CO_2/brine from CO_2 geological storage formations. This paper intends to describe the mathematical basis of the linking function approach and test its application to a relatively simple hypothetical case study. The application of LFSM, with chemical scaling function being a particular case, for the evaluation of risk of CO_2 site on shallow groundwater is given in Carroll et al. (2014).

2 Method

2.1 Assumptions and formulations

We consider a deterministic high-fidelity model (HFM) calculating a scalar output $Y_{HFM} = f\,(\boldsymbol{x}_{HFM})$ for a set of input parameters $\boldsymbol{x}_{HFM} = (x_1, \ldots, x_n)$. We assume that the physical system considered by the HFM can also be simulated with lower-fidelity models (LFMs), each of which representing a simplification of the HFM. Differently from previous works (*Kennedy and O'Hagan*, 2000; *Quian et al*., 2006 and 2008; *Huang et al*., 2006; *Goh et al*., 2012), we do not require a priori ranking of the lower-fidelity models in terms of fidelity, nor do we assume a hierarchical framework. We only assume that the HFM is more computationally demanding than the LFMs because it takes into account a larger number of parameters and processes describing the physical system. On the other hand, the lower-fidelity models provide only a partial representation of the complexity of the system, but their execution times are shorter than the HFM. Different approaches can be taken for simplifying the HFM. For example, the lower-fidelity models may have a coarser spatial discretization of the model domain with respect to the HFM, or a lower dimensional representation (i.e., two-dimensional instead of three-dimensional). In other cases, the lower-fidelity models may be less accurate because they do not consider heterogeneity, or because some of the physical or chemical processes that are simulated by the HFM are not taken into account. Furthermore, simplifying assumptions about the conceptual model of the physical system may be made, allowing the use of analytical solutions in the lower-fidelity models rather than complex numerical solutions provided by the HFM.

By defining the lower-fidelity as simplifications of the HFM, we made the fundamental assumption that the HFM and the lower-fidelity models share some basic features and therefore are correlated in some way (*Kennedy and O'Hagan,* 2000). With this assumption, we can identify the input parameters of the lower-fidelity models as subsets of the set of input parameters of the HFM ($\boldsymbol{x}_{HFM}$). For the case with two lower-fidelity models, we can then write the output from the first lower-fidelity model (LFM-1) as $Y_{LFM\text{-}1} = f(\boldsymbol{x}_{LFM\text{-}1})$, where $\boldsymbol{x}_{LFM\text{-}1}$ is a subset of $\boldsymbol{x}_{HFM}$, and the output from the second lower-fidelity model (LFM-2) as $Y_{LFM\text{-}2} = f(\boldsymbol{x}_{LFM\text{-}2})$, where $\boldsymbol{x}_{LFM\text{-}2}$ is another subset. We propose that the relationship between the outputs from the HFM and from the two lower-fidelity models can be represented by a mathematical function, and therefore the output from the HFM can then be written as:

$$Y_{HFM} = g(Y_{LFM-1}, Y_{LFM-2}, \boldsymbol{\beta}) + \varepsilon \tag{1}$$

where g is a mathematical function, hereafter referred to as the "*linking function*", $\boldsymbol{\beta}$ is a vector of unknown parameters of the linking function, and ε is a regression error term. Equation (1) is the core of the LFSM approach. The linking function represents a surrogate model that "links" outputs from models with different levels of fidelity, and formally addresses their discrepancies.

So far, we considered a case with only three levels of fidelity (HFM, LFM-1 and LFM-2). However, we can easily extend our approach to multiple levels of fidelity. Suppose the HFM can be simplified with a number *k* of lower-fidelity models Y_{LFM-i} ($i = 1, \ldots, k$), each sharing some of the input parameters and simulated processes of the high-fidelity model. Extending Equation (1) to such a case, the output from the HFM can then be defined as follows:

$$Y_{HFM} = g(Y_{LFM-i}, \boldsymbol{\beta}) + \varepsilon \quad (i = 1, \ldots, k) \tag{2}$$

A special case is when only two levels of fidelity are considered (i.e., HFM and LFM-1). In this case, the proposed methodology can be seen as similar to another surrogate modeling approach, known as Space Mapping (*Bandler et al*., 1994, *Robinson et al*. 2006, 2008).

2.2 Shape of the linking function

The linking function can assume different forms, which must be defined on a case-by-case basis. In very simple cases, the shape of the linking function might be obvious and be defined on the basis of the physics of the problem. However, for more complex problems where there is no obvious relationship, an empirical relationship must be adopted. In these situations, the implementation of the linking function is analogous to the development of a response surface representing the relationship between lower-fidelity models outputs and the correspondent outputs from the high-fidelity model. In the field of response surface surrogate modeling, different mathematical approaches have been applied to approximate the relationships between input parameters, also known as explanatory variables, and the original model output. The most popular approaches include polynomials, kriging, artificial neural networks, radial basis functions, and multivariate adaptive regression splines. The same mathematical functions can be used as linking functions, and we refer to several comparative studies (e.g., *Giunta et al.*, 1998; *Simpson et al.*, 2001; *Stander et al.*, 2004; *Fang et al.*, 2005; *Forsberg and Nilsson*, 2005; *Wang and Shan*, 2007; *Zhao and Xue*, 2010; *Razavi et al.*, 2012b) for the details of each method and discussions on their advantages and disadvantages.

In this work we focus on polynomials, which can be very flexible and take a wide variety of functional forms. Assuming the linking function g in Equation (2) is a polynomial of degree n, then it can be written as:

$$g(Y_{LFM-i}, \boldsymbol{\beta}) = \beta_0 + \sum_i \beta_i \, Y_{LMF-i} + \sum_i \sum_{j>i} \beta_{ij} \, Y_{LFM-i} Y_{LFM-j} + \sum_i \beta_{ii} \, {Y_{LFM-i}}^2 + \sum_i \sum_{j>i} \sum_{k>j} \beta_{ijk} \, Y_{LFM-i} Y_{LFM-j} Y_{LFM-k} + \cdots + \sum_i \beta_{ii\ldots i} \, {Y_{LFM-i}}^n \tag{3}$$

The coefficients $\boldsymbol{\beta}$ of the polynomial are determined through the least-squares solution of the equation $\mathbf{G}\boldsymbol{\beta} = \boldsymbol{Y}_{HFM}$, where $\mathbf{G}$ is a matrix operator, and $\boldsymbol{Y}_{HFM}$ is a vector of outputs determined from a number m of runs of the HFM (see Section 2.3). The maximum likelihood estimates of the coefficients are then defined as $\boldsymbol{\beta} = (\mathbf{G}^{\mathrm{T}}\mathbf{G})^{-1}\mathbf{G}^{\mathrm{T}}\boldsymbol{Y}_{HFM}$. If we consider for simplicity three levels of fidelity – one high-fidelity model and two lower-fidelity models (LFM-1 and LFM-2) – and we assume that the linking function can be written in the form of a 2nd order polynomial (n = 2), the matrix operator $\mathbf{G}$ is defined as:

$$\boldsymbol{G} = \begin{bmatrix} 1 & Y_{\mathrm{LFM-1}_1} & Y_{\mathrm{LFM-2}_1} & Y_{\mathrm{LFM-1}_1}Y_{\mathrm{LFM-2}_1} & {Y_{\mathrm{LFM-1}_1}}^2 & {Y_{\mathrm{LFM-2}_1}}^2 \\ 1 & Y_{\mathrm{LFM-1}_2} & Y_{\mathrm{LFM-2}_2} & Y_{\mathrm{LFM-1}_2}Y_{\mathrm{LFM-2}_2} & {Y_{\mathrm{LFM-1}_2}}^2 & {Y_{\mathrm{LFM-2}_2}}^2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & Y_{\mathrm{LFM-1}_m} & Y_{\mathrm{LFM-2}_m} & Y_{\mathrm{LFM-1}_m}Y_{\mathrm{LFM-2}_m} & {Y_{\mathrm{LFM-1}_m}}^2 & {Y_{\mathrm{LFM-2}_m}}^2 \end{bmatrix} \quad (4)$$

where $Y_{LFM\text{-}1_t}$ and $Y_{LFM\text{-}2_t}$, with $t = (1, \dots, m)$, are the outputs from the t^{th}-run of the first and second lower-fidelity model, respectively. Since the n^{th}-order polynomial approximation of a certain function can be seen as a Taylor Series expansion of the function truncated after n+1 terms (Box and Draper, 1987), higher-order polynomials (more expansion terms) can provide a more accurate approximation. However, since the number of input parameters is usually large in most of the engineering and environmental applications, the use of higher-order polynomials ($n >$ 2) as response surface surrogates is often infeasible [*Forrester et al*, 2007; *Razavi et al*., 2012a]. This is because the minimum number of runs (m_{min}) required for the estimation of the coefficients **β**, which is a function of both the number of parameters and the order n, may become prohibitively large for high-dimensional problems. For a D-dimensional parameter input space, m_{min} is given by:

$$m_{min} = \frac{(n + D)!}{n!\,D!} \quad (5)$$

For instance, at least 1001 runs of the original model are required to estimate the coefficients of a 4^{th}-order polynomial for a system with ten explanatory variables (D = 10). The issue of dimensionality is not expected to be a factor in the application of polynomials as linking functions. This is because the number of possible lower-fidelity models, which in the proposed LFSM framework represents the variable D in Equation (5), is expected to be much lower than the number of input parameters of the original HFM. For this reason, with respect to the application of polynomials, the LFSM approach has two immediate advantages over traditional response surface modeling approach. The first is that fewer HFM runs are required for estimating the coefficients of the polynomial. The second advantage is that polynomial linking functions of higher order, which may provide a better representation of the relationship between the models outputs, can be adopted with a relatively small number of additional runs of the HFM. In the case of a system with three levels of fidelity, for example, only 9 additional runs are necessary to develop a 4th order polynomial linking function instead of a 2nd order polynomial. Nevertheless, we emphasize the use of standard model selection criteria used in regression analysis (e.g., *Akaike*, 1973; *Schwarz*, 1978), to avoid the problem of over fitting or the risk of developing an excessively complex model that may yield a poor generalization.

2.3 Overview of the procedure

The following steps represent a guide through the implementation of the proposed LFSM approach for a general case study. A schematization of the procedure for three models with different levels of fidelity is shown in Figure 1.

Step 1. Implement the high-fidelity model (HFM) that provides the most realistic representation of the system of interest by taking into account the highest number of processes and parameters controlling the system.

Step 2. Identify strategies to simplify the HFM, and apply them to the implementation of the lower-fidelity models (*LFM-i*). Different approaches may be taken for the simplification of the

HFM such as simplification of the conceptual model, coarsening of the spatial or temporal discretization, parameter upscaling, and exclusion of certain physical or chemical processes. However, it is important that the HFM and the LFM-*i* are somehow correlated, meaning that input parameters of the lower-fidelity models correspond to subsets of the input parameters of the HFM. If, for instance, x_i (i = 1,..., n) is the set of n-input parameters of the HFM, then the subset x_j ($j = r, \ldots, p$), where $r \geq 1$ and $p < n$, may represent the input parameters of the lower-fidelity model LFM-1. If a second lower-fidelity model LFM-2 is implemented, which may take into account the processes and parameters omitted in LFM-1, then its input parameters are defined in another subset x_k ($k = s, \ldots, n$). If $s < p$, as shown in Figure 1, then the LFM-2 model shares some of the input parameters with both the HFM and the LFM-1. However, the LFSM approach can also be applied when $s \geq p$.

Step 3. Generate a sample m of input parameters for the HFM. From the generated input parameters sets, extract the subsets corresponding to each LFM-*i*. Sample generation should be made according to one of the design of experiments (DoE) methods to minimize the number of HFM model evaluations, while maximizing our understanding of the model behavior. There is a large variety of space-filling DoE strategies in the literature, including fractional factorial sampling, Latin hypercube sampling, and different strategies based on sequences of quasi-random numbers. Details on DoE methods can be found, for example, in *Saltelli et al.* [2008].

Step 4. Perform m runs of the HFM and the LFM-*i* models with the generated samples of input parameters to calculate the vectors of outputs $\boldsymbol{Y}_{HFM} = (Y_{HFM,1}, \ldots, Y_{HFM,m})$ and $\boldsymbol{Y}_{\mathrm{LFM\text{-}i}} = (Y_{LFM\text{-}i,1}, \ldots, Y_{LFM\text{-}i,m})$, for each of the implemented models.

Step 5. Use the generated numerical data $\boldsymbol{Y}_{HFM}$ and $\boldsymbol{Y}_{LFM\text{-}i}$ to identify a mathematical function (i.e., the linking function), representing the best match between outputs from the different models. In practice, this step consists of a regression analysis in which Y_{HFM} is the dependent variable and the outputs from the lower-fidelity models $Y_{LFM\text{-}i}$ are the independent variables. The shape and the coefficients of the linking function can be estimated with the least-square-regression method or other methods (see Section 2.2). Once the linking function has been identified, the linking function surrogate model can be used to emulate the output from the HFM by: 1) running the lower-fidelity models with a set of parameters of interest; 2) use lower-fidelity responses as inputs in the linking function and approximate the response of the HFM.

Before illustrating the proposed method with a numerical example, we provide a few more comments on the proposed approach. The application of a linking function model can be very advantageous especially for certain types of numerical investigations, such as engineering design optimization or global sensitivity analysis, which require multiple runs of a potentially slow and numerically unstable HFM. By addressing the discrepancies between the HFM and the lower-fidelity models, the linking function can retain the level of realism and detailed information associated with the HFM, while at the same time avoiding the long computational times usually associated with running such models. However, an important factor to consider is the numerical efficiency of the lower-fidelity models. It is obvious that the LFSM approach is attractive only if the sum of execution times of the lower-fidelity models is appreciably lower than the time required for running the HFM. In this regard, it is noteworthy that with our approach the lower-fidelity models can be substituted by other types of surrogate models such as previously developed response surface models or ROMs. With more computational effort, response surface surrogates of the physically-based lower-fidelity models can even be developed at the same time as the linking function. This will translate to an even faster emulation of the HFM response.

The LFSM approach is based on the assumption that there is a correlation, represented by the linking function, between the outputs from the lower-fidelity models and those from the HFM. Clearly, if the goodness-of-fit of the linking function is poor, or is acceptable only in selected sectors of the input parameter space, the emulated outputs will not be accurate. Finding an accurate linking function may become an issue when model outputs for a given set of input parameters are very different from the outputs for a slightly different set. To some extent, this issue can be solved by changing the strategy used for simplifying the HFM because it may be the result of the oversimplification of the HFM, which causes the development of lower-fidelity models that are not sufficiently representative of the system of interest. However, increasing the level of fidelity of the lower-fidelity models may have the effect of increasing their execution times, which inevitably reduces the computational efficiency of the approach.

Although the development of the linking function requires some computational cost, given by the time to collect the necessary data from the model runs, this cost is expected to be less than that required for the development of a response surface surrogate for the same system of interest. This is because the number of HFM runs required for the development of a robust response surface is a function of the number of input parameters. Conversely, the number of HFM runs required for developing a robust linking function depends on the number of lower-fidelity models considered, which will always be less than the number of input parameters in the HFM. The limitation of the number of HFM runs provides another advantage. It allows the HFM to consider a higher level of complexity than would be feasible in the development of a response surface model.

3 Example of application to CO_2 storage risk assessment

The LFSM approach was employed to build the reduced order models (ROM) that predict the impact of a hypothetical CO_2 and brine leakage on groundwater quality in an hypothetical aquifer with hydrological and hydrochemical properties similar to those of the High Plains aquifer (USA). To define the impact of leakage on groundwater quality, we performed numerical simulation of multiphase flow and reactive transport to estimate concentrations of certain chemical species. The impact of leakage was then measured by calculating the following metrics:

- Volume of aquifer reaching pH < 6.5;
- Volume of aquifer reaching TDS > 500 mg/L.
- Volume of aquifer reaching concentrations of As > 1.33×10^{-7} mol/kg;
- Volume of aquifer reaching concentrations of Cd > 4.05×10^{-8} mol/kg;
- Volume of aquifer reaching concentrations of Pb > 7.24×10^{-8} mol/kg;

Statistical analyses were conducted on the groundwater concentration data collected in a 2010 U.S. Geological Survey (USGS) groundwater survey of 30 wells within the High Plains aquifer and thresholds for eastimating the volume of impacted aquifer (e.g. 1.33×10^{-7} mol/kg for As) were calculated as the 95%-confidence, 95%-coverage tolerance from the data.

3.1 Problem statement and hydrogeological setting

We consider a two-dimensional cross-section of a hypothetical aquifer of length equal to 10,000 m and thickness equal to 240 m. The lithological characterization of the aquifer is based on the lithological descriptions of 48 wells located in Haskell County, in South West Kansas. The source of these data is the Water Well Completion Records (WWC5) Database (Kansas Geological Survey, 2012). Lithological descriptions of the well logs include different types of unconsolidated sediments with a highly heterogeneous granulometric distribution, typical of a fluvial depositional environment. For simplicity, the original lithological descriptions were classified into two hydrostratigraphic units on the basis of grain size (coarse/fine) and permeability (high/low). The lithologies included in each of these units are provided in Table 1. The aquifer is assumed to be confined, and the mean groundwater flow is from east to west with a hydraulic gradient of 0.003. Aquifer thickness is uniform and equal to 240 m, which corresponds to the average thickness of the High Plain Aquifer in Heskell County.

The distribution of the two hydrostratigraphic units was simulated with the T-PROGS approach (Carle and Fogg, 1996 and 1997; Carle et al., 1998; Carle, 1999) based on the transition probabilities between different categories and on a single Markov Chain equation in each direction. Transition probabilities are defined as the probability that a certain category *j* occurs at the location **u**+**h** conditioned to the occurrence of another category *i* at the location **u**. Here **u** and **h** are a location and a movement vector, respectively. One advantage of this methodology is the increased realism of the simulations, making it thereby possible to account for observable geological features such as mean lengths and juxtapositional tendencies. T-PROGS simulations of the aquifer heterogeneity were conducted with mean lengths and volumetric proportions for the two different hydrostratigraphic estimated from the analysis of their spatial distributions in the 48 wells. Values for these two parameters are given in Table 1. The spatial distribution of the two hydrostratigraphic units corresponds to one unconditional realization of the T-PROGS geostatistical model. The interpolation grid is composed of rectangular cells with constant dimensions equal to 100 m and 5 m in the x-direction and z-direction, respectively (Figure 2). In the implementation of the multiphase and transport models described in the next sections, different hydrological parameters (i.e., permeability, porosity, etc.) were assigned to each unit.

Leakage of CO_2 and brine from a wellbore is simulated by assuming a point source at the point of coordinates (2200 m, -190 m) and a duration of 200 years with variable leakage rates. The maximum values of these leakage rates are plotted as a function of simulation time in Figure 3. These rates represent a hypothetical leakage pathway related to a deep leaky well connecting a deep geologic reservoir for CO_2 storage with a shallow groundwater resource. We assumed that leakage is driven by reservoir over pressure and CO_2 and brine saturations. These parameters were used as input in a wellbore leakage model based on multiphase and non-isothermal flow simulations (Jordan et al., 2013) to calculate the flux into the aquifer and the leakage rate over time. The CO_2 rates sharply increase during the initial 5 years and then oscillate, with variations ranging from 0.039 kg/s to 0.046 kg/s. The brine leakage rates are more stable, with very little variation around an average of 0.012 kg/s.

3.2 High and lower fidelity models

Leakage of CO_2 and brine in the hypothetical aquifer was modelled with a 2-D high fidelity model (HFM) considering a comprehensive set physical and chemical factors, as well as with two lower-fidelity models (LFM-1 and LFM2), which take into account only some of the parameters and processes considered by the HFM. For all models, multiphase flow and reactive transport were simulated with the finite-volume code TOUGHREACT 2.0 (Xu et al., 2011).

The first lower fidelity model (LFM-1) is a simple model considering 1-D flow parallel to the average flow direction in the hypothetical aquifer. The simulation domain is10,000 m in X direction which is discretized into 1000 grid blocks and 1 m in Y and Z direction without discretization. The aquifer is considered homogenous, and a hydraulic gradient equal to 0.003 is applied by fixing the pressure at grid blocks on the left and right boundaries. Chemical reactions are considered in the model including aqueous complexation, mineral dissolution/precipitation, cation exchange and adsorption/desorption via surface complexation. Details of these reactions are given in Bianchi et al. (2013). In this model, the dissolution of calcite and surface protonation reactions are the main pH buffering processes. Surface complexation reactions on goethite, illite, kaolinite and smectite are the dominant reactions that control the release of As, Pb and Cd.

The second lower fidelity (LFM-2) model simulates 2-D flow and solute transport and assumes a heterogeneous distribution of permeability in the hypothetical aquifer (Figure 2). The simulation domain is 10,000 m in X direction, 240 m in Z direction and 1000 m in Y direction. However, the domain is discretized only in the X and Z directions Unlike LFM-1, LFM-2 does not consider chemical processes i.e. the chemical species are treated as conservative species. Specified hydraulic head boundary conditions were imposed at the left and right boundaries, while no-flow boundary conditions were applied at the top and bottom of the domain. A preliminary gravity equilibration run, without CO_2 and brine injection, was run long enough to establish quasi-steady-state initial conditions, for the CO_2 and brine leakage simulations. These were conducted at constant temperature (17^{o}C). Simulated plumes for the considered chemical species are shown in Figures 4. These results are representative of a base-case run used to understand the system behavior and to manually test the sensitivity of the model to the different input parameters. Input parameters for this base-case scenario are presented in Table 2. After 200 years of continuous release of brine and CO_2 from the leakage point, the area with lowered pH values and the plumes of three considered metals (As, Pb, and Cd) moved about 7 km downgradient. As expected, the major role in determining the shape of these plumes is played by the heterogeneous distribution of the two hydrostratigraphic units, with the plume following preferential flow paths according to the distribution of the highest permeable unit.

The high fidelity model (HFM) considers multi-phase flow and transport in a heterogeneous system and geochemical reactions. The model setup, hydrogeological parameterization, and leakage functions of CO_2 and brine are the same as in the lower-fidelity model LFM-2. The HFM also incorporates all the chemical reactions considered by LFM-1. Because of its complexity, this HFM is expected to provide the most accurate representation of the natural system, taking into account uncertainties in flow, transport, and chemical processes. Figure 5 shows the plumes of pH, As, Pb, and Cd for a base-case simulation similar to that considered for LFM-1 and LFM-2.

3.3 Development of the linking function surrogate model and results

Multiple runs of the HFM and of the two lower-fidelity models LFM-1 and LFM-2 were performed to collect the data required for developing the linking function. Initially, 450 sample points in the HFM parameter space were generated using a quasi-random sequence algorithm (LPτ , Sobol et al. 1992). Correspondent input parameters for the lower-fidelity models were then extracted from the generated sets, according to the previously described procedure. Details on the considered input parameters and on their ranges are presented in Table 3.

For each set of input parameters, we ran the HFM and the two lower-fidelity models to estimate three different predictions (i.e., one from each model) of the impact of CO2 and brine leakage for 20 simulated time periods (one every 10 years up to 200 years). At the end of these numerical simulations, three vectors of output values were estimated and used in the least-squares regression analysis to estimate the coefficients of the polynomial representing the linking function between the HFM and LFM-1 and LFM-2 (Equation 3). For all the output variables and for all the simulation times, a second order polynomial was found to provide a sufficiently accurate match between the simple and complex model outputs. An example of the shape of this polynomial function is shown in Figure 6. The goodness of fit for the developed linking functions was analyzed by calculating the coefficient of determination (R^2). Taking into account all the linking functions, the calculated R^2 values are between 0.635 and 0.998, with the majority of values higher than 0.800. A better accuracy in terms of R^2 values can be obviously obtained by augmenting the order of the polynomial linking functions, but this can also increase the risk of overfitting and, consequentially, compromise the predictability power of the developed LFSM. In general the highest R^2 values, indicating higher accuracy, are calculated for the linking functions that predict the volume of TDS > 500 mg/l (R^2 between 0.970 and 0.987). For the linking functions considering pH, the R^2 values ranges between 0.822 and 0.944. Relatively less accuracy is associated with the linking functions for estimating the volume of aquifer contaminated with As (R^2 between 0.911 and 0.723), Pb (0.725 - 0.638) and Cd (0.753 - 0.635)We also built scatter plots of the responses estimated with the linking function and those of the complex model (Figure 7). In general, the cloud of points is distributed along a y = x line, showing the accuracy of the fitting. The highest accuracy is for smaller simulations times.

3.4 Application of the linking function surrogate models in NRAP

The development of ROMs generally relies on conducting a number of high-fidelity numerical simulations that consider all relevant flow, transport, and chemical processes that could potentially have an impact on CO_2 and brine leakage into groundwater. These high-fidelity simulations are then used to "train" simpler ROMs (e.g., look-up tables, functional relationships) that sufficiently represent their outputs for a wide range of uncertain input parameters. To overcome the problem of running a very complex and extremely computationally demanding model considering all the parameters and processes that are relevant to brine and CO_2 leakage in shallow aquifers, within NRAP we make an attempt to represent the complex hydrogeological and geochemical conditions in a heterogeneous aquifer by using two separate lower-fidelity models. The outputs from these lower-fidelity models are then used as input for the linking functions developed in this work. In particular, one lower-fidelity model is represented by a ROM that estimates the volume impacted by CO_2 and brine leakage by taking into account heterogeneous flow and transport conditions in the presence of multiple leakage wells. This ROM, developed by Lawrence Livermore National Laboratory (Carroll, et al., 2013) and referred to as hydrological ROM, considers uncertainties related to flow, transport, and leakage parameters, but it has a simplified representation of the chemical reactions induced by leakage. Complex chemical reactions are instead considered in detail by a second ROM, which, on the other hand, does not include parameters for a more accurate representation of the hydrological complexity. In particular, the input parameters of this ROM do not include properties defining aquifer heterogeneity. This ROM which was developed by Lawrence Berkeley National Laboratory, and is here referred to as geochemical ROM, allows to define uncertainties related to chemical parameters and reactions. In the context of the previously described linking function development, the geochemical ROM is comparable to the lower-fidelity model LFM-1, while the hydrological ROM is comparable to the lower- fidelity model LFM-2. Moreover, the similarity between the two previously described lower fidelity models and the two ROMs is also determined by the fact that all models were developed on the basis of physical and chemical parameters values that are consistent with the characteristics of the High Plains Aquifer (Becker et al., 2002).

Clearly, neither the hydrological ROM nor the geochemical ROM can separately provide an accurate prediction of the risk profile, because of the simplifications inherent in these models. Therefore, we used the developed LFSM to link the outputs from the two ROMs as described in the workflow shown in Figure 8. In practice, once the outputs from the hydrological and geochemical ROMs for a particular objective variable (e.g., TDS, pH, As, Cd, or Pb) and simulation time are obtained, these are directly used as input in the corresponding linking functions to estimate the final volume of aquifer impacted by CO_2 and brine leakage. This final volume is expected to represent a reasonable approximation of the output from a time consuming and computationally expensive computer model. In theory, a complex 3-D model should be ideally implemented and run multiple times to estimate the HFM outputs for the development of the linking functions to link the outputs from the two ROMs (Figure 1). However, due to the complexity of the systems considered in NRAP, the development of such high-fidelity numerical model that incorporates 3-D heterogeneous flow and reactive transport is very challenging and too computationally demanding with the currently available numerical codes. Therefore, for now, we have made the assumption that the linking functions developed to emulate the outputs from a complex 2-D model can be extrapolated in 3-D and provide reliable estimates of impacted aquifer volumes. However, once computational power will be available, future research should dedicate to test the reliability of the 2-D assumption by comparing the outputs from the developed LFSMs with the corresponding outputs from a complex 3-D model.

4 Conclusions

We present a new surrogate modeling approach, named Linking Function Surrogate Modeling (LFSM), which is based on the simplification of a computationally expensive high-fidelity model into computationally less expensive and simpler models, followed by the development of a mathematical function that links their outputsto emulate the output from the high-fidelity model.. When the sum of execution times of the lower-fidelity models is less than the execution time of the high-fidelity model, this approach can significantly reduce the computational cost without jeopardizing the accuracy of the results. In comparison with other surrogate modeling methods, the main advantage of the proposed approach is that it can manage problems where the number of input parameters of lower-fidelity models is different from that in the high-fidelity model.

The proposed approach was applied to the development of reduced order models that estimate the impact of CO_2 and brine leakage on groundwater quality in a heterogeneous shallow aquifer. A computationally expensive high-fidelity multiphase flow and reactive transport model (HFM) was simplified into two lower-fidelity models, LFM-1 and LFM-2, each taking into account a subset of the processes simulated in the HFM. In particular, LFM-1 is a simple 1-D model with homogenous flow field, but it takes into account several chemical reactions. On the other hand LFM-2 is 2-D model considering aquifer heterogeneity but no reactions. We showed that outputs from the HFM can be emulated with satisfactory accuracy with the proposed linking function surrogate modelling approach. For all the model responses considered, the linking functions are represented by 2nd order polynomials. These functions were developed with a limited computational cost through a least-squares regression analysis in which the outputs from the LFM-1 and LFM-2 are used as independent variables to fit the outputs from HFM.

Within NRAP, the developed linking functions are applied to link the output from two ROMs. These two ROMs are in fact similar in terms of input parameters and considered processes to the lower-fidelity models LFM-1 and LFM-2 used for the testing the proposed linking function approach. The first ROM (hydrological ROM) was in fact derived from a low-fidelity model that accounts for the heterogeneous flow and transport conditions in the presence of multiple leakage wells, which considered uncertainties related to flow, transport, and leakage parameters, but has a simplified representation of chemical reactions. The second ROM (geochemical ROM) was obtained from models that feature greatly simplified flow and transport conditions, but allow for a more complex representation of relevant geochemical reactions. This ROM deals with uncertainties related to chemical parameters and reactions. The proposed linking function approach allows to combine the outputs from these two ROMs to provide estimations of of ivolume of aquifer impacted by CO_2 and brine leakagewithout the need of performing multiple evaluations of a complex and computationally infeasible high-fidelity model.

Acknowledgments. This work was funded by the Assistant Secretary for Fossil Energy, National Energy Technology Laboratory, National Risk Assessment Partnership, of the US Department of Energy under Contract No. DEAC02-05CH11231. The Guest Editor Susan Carrol and two anonymous reviewers are grateful acknowledged for their comments, which helped improving the final version of the manuscript.

References

Akaike, H. (1973), Information theory as an extension of the maximum likelihood principle. Pages 267-281 *in* B. N. Petrov and F. Csaki, editors. Second international symposium on information theory. Akademiai Kiado, Budapest, Hungary.

Alexandrov, N.M., R. M. Lewis, C. R. Gumbert, L. L. Green, and P. A. Newman (2001), Approximation and model management in aerodynamic optimization with variable-fidelity models, Journal of Aircraft, 38 (6), 1093–1101.

Annoni, P., R. Brüggemann, and A. Saltelli (2011), Partial order investigation of multiple indicator systems using variance-based sensitivity analysis, Environmental Modelling & Software, 26(7), 950-958, ISSN 1364-8152, 10.1016/j.envsoft.2011.01.008.

Bandler, J., Biernacki, R., Chen, S., Grobelny, P., and Hemmers, R. (1994), Space Mapping Technique for Electromagnetic Optimization, IEEE Transactions on Microwave Theory and Techniques, 42(12), 2536–2544, doi:10.1109/22.339794.

Baú, D., and A. S. Mayer (2006), Stochastic Management of Pump-and-Treat Strategies Using Surrogate Functions, Adv. Water Resources, 29(12), 1901-1917.

Berci, M., P. H. Gaskell, R. W. Hewson, and V. V. Toporov (2011), Multifidelity metamodel building as a route to aeroelastic optimization of flexible wings, Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 225, 2115-2137, doi: 10.1177/0954406211403549.

Becker, M. F, B. W. Bruce, L. M. Pope, and W. J. Andrews, 2002, Ground-Water Quality in the Central High Plains Aquifer, Colorado, Kansas, New Mexico, Oklahoma, and Texas, 1999. U.S. Geological Survey, Water-Resources Investigations Report, 02-4112.

Bianchi, M., Zheng, L., Spycher, N., Birkholzer, J. *Reduced Order Models for Prediction of Groundwater Quality Impacts from CO_2 and Brine Leakage: Application to the High Plains Aquifer,* NRAP-TRS-III-xxx-2013; NRAP Technical Report Series; U.S. Department Energy, National Energy Technology Laboratory: Morgantown, WV, 2013; p52.

Bliznyuk, N., D. Ruppert, C. Shoemaker, R. Regis, S. Wild, and P. Mugunthan (2008), Bayesian calibration and uncertainty analysis for computationally expensive models using optimization and radial basis function approximation, J. Comput. Graphical Stat., 17(2), 270–294.

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

Box, G. E. P., and N. R. Draper (1987), Empirical Model-Building and Response Surfaces, John Wiley & Sons, New York, NY.

Carroll, S. A., Mansoor, K., and Sun, Y. Second Generation Reduced Order Model for Calculating Groundwater Impacts as a Function of pH, Total Dissolved Solids, and Trace Metal Concentration; NRAP-TRS-III-2013; NRAP Technical Report Series; U.S. Department of Energy, National Energy Technology Laboratory: Morgantown, WV, 2013.

Carroll, S. A., Bianchi, M., Mansoor, K., Zheng, L., Sun, Y., Spycher, N., and Birkholtzer, J. (2014). "Reduced-Order Model for Estimating Impacts from CO2 Storage Leakage to Alluvium Aquifers: Third-Generation Combines Physical and Chemical Processes." NRAP

Technical Report Series; U.S. Department of Energy, National Energy Technology Laboratory: Morgantown, WV, 2014; p 39.

Cui, T., Fox, C., and M. J. O'Sullivan (2011), Bayesian calibration of a large scale geothermal reservoir model by a new adaptive delayed acceptance Metropolis Hastings algorithm, Water Resour. Res., 47, W10521, doi:10.1029/2010WR010352.

Dai, Z., E. Keating, D. Bacon, H. Viswanathan, P. Stauffer, A. Jordan and R. Pawar 2014. Probabilistic evaluation of shallow groundwater resources at a hypothetical carbon sequestration site. Scientific Reports 4: 4006.

Deutsch, C., and A. Journel (1998), GSLIB Geostatistical Software Library and User's Guide, 2nd edition, Oxford Univ. Press, N. Y.

Doherty, J. and S. Christensen (2012), Use of paired simple and complex models to reduce predictive bias and quantify uncertainty, Water Resour. Res., 47, W12534, doi:10.1029/2011WR010763.

Dyn, N., D. Levin, and S. Rippa (1986), Numerical procedures for surface fitting of scattered data by radial basis function, SIAM J Sci Statist Comput, 7(1), 639–659.

Eldred, M. S., A. A. Giunta, S. S. Collis (2004), Second-order corrections for surrogate-based optimization with model hierarchies. *In*: 10th AIAA/ISSMO multidisciplinary analysis and optimization conference, Albany, New York, 30–31 August 2004.

Fang, H., M. Rais-Rohani, Z. Liu, and M. F. Horstemeyer (2005), A comparative study of metamodeling methods for multiobjective crashworthiness optimization, Comput. Struct., 83(25–26), 2121–2136.

Fen, C. S., C. C.Chan, and H. C.Cheng (2009), Assessing a response surface-based optimization approach for soil vapor extraction system design, J. Water Resour. Plann. Manage., 135(3), 198–207.

Forrester, A. I. J., A. Sobester, and A. J. Keane (2007), Multi-fidelity optimization via surrogate modelling, Proc. R. Soc. A, 463(2088), 3251–3269.

Forsberg, J., and L. Nilsson (2005), On polynomial response surfaces and kriging for use in structural optimization of crashworthiness, Struct. Multidisc. Optim., 29(3), 232–243.

Forrester, A. I. J. (2010), Black-box calibration for complex-system simulation, Phil. Trans. R. Soc. A, 368, 3567–3579, doi:10.1098/rsta.2010.0051.

Forrester A. I. J., and A. J. Keane (2009), Recent advances in surrogate-based optimization, Prog. Aerosp. Sci., 45(1–3), 50–79.

Gano, S. E., J. E. Renaud, J. D. Martin, and T. W. Simpson (2006), Update strategies for kriging models used in variable fidelity optimization, Struct. Multidiscip. Optim., 32(4), 287–298.

Giunta, A., L. T. Watson, and J. Koehler (1998), A comparison of approximation modeling techniques: polynomial versus interpolating models. In: Proceedings of the 7th AIAA/USAF/NASA/ISSMO symposium on multidisciplinary analysis & optimization, No. AIAA-98–4758, St. Louis, MO, 1998.

Goh, J., D. Bingham, J. P. Holloway, M. J. Grosskopf, F. W. Doss, E. Rutter, and C. C. Kuranz (2012), Computer Model Calibration Using Outputs From Multi Fidelity Simulators, American Physical Society, 53rd Annual Meeting of the APS Division of Plasma Physics, November 14-18, 2011, abstract #PP9.132.

Harbaugh, A. W., E. R. Banta, M. C. Hill, and M. G. McDonald (2000), MODFLOW-2000, the U.S.Geological Survey Modular Ground-Water Model—User Guide to Modularization Concepts and the Ground-Water Flow Processes. U.S. Geological Survey Open-File Report 00-92. Reston, Virginia: U.S. Geological Survey.

Hemker, T., K. R. Fowler, M. W. Farthing, and O. vonStryk (2008), A mixed-integer simulation-based optimization approach with surrogate functions in water resources management, Optim. Eng., 9(4), 341–360.

Huang, D., T. T. Allen, W. I. Notz, and R. A. Miller (2006), Sequential kriging optimization using multiple-fidelity evaluations, Struct. Multidisc. Optim., 32, 369–382, DOI:10.1007/s00158-005-0587-0

Jordan, A. B.; Stauffer, P. H.; Harp, D. R.; Carey, J. W.; Pawar, R. J. A Method for Predicting CO2 and Brine Leakage from Geologic Sequestration along Cemented Wellbores in System-Level Models. Los Alamos National Laboratory Rep. LA-UR-13-29243; 2013.

Jordan, A. B., P. H. Stauffer, D. Harp, J. W. Carey and R. J. Pawar 2015. A response surface model to predict CO2 and brine leakage along cemented wellbores. International Journal of Greenhouse Gas Control 33: 27-39.

Jones, D. (2001), A Taxonomy of Global Optimization Methods Based on Response Surfaces, Journal of Global Optimization, 21(4), 345–383, doi:10.1023/A:1012771025575.

Kansas Geological Survey (2012). Water Well Completion Records (WWC5) Database. Available at www.kgs.ku.edu/Magellan/WaterWell/index.html.

Kennedy, M. C., and A. O'Hagan (2000), Predicting the output from a complex computer code when fast approximations are available, Biometrika, 87(1), 1–13.

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

Kim, H. S., M. Koc, and J. Ni (2007), A hybrid multi-fidelity approach to the optimal design of warm forming processes using a knowledge-based artificial neural network, International Journal of Machine Tools & Manufacture, 47 (2), 211–222.

Kleijnen, J. P. C. (2009), Kriging metamodeling in simulation: A review, Eur. J. Oper. Res., 192(3), 707–716.

Koch, P. N., T. W. Simpson, J. K. Allen, and F. Mistree (1999), Statistical Approximations for Multidisciplinary Optimization: The Problem of Size, Special Multidisciplinary Design Optimization Issue of Journal of Aircraft, 36(1), 275–286.

Kourakos, G., and A. Mantoglou (2009), Pumping optimization of coastal aquifers based on evolutionary algorithms and surrogate modular neural network models, Adv. Water Resour., 32(4), 507–521.

Kourakos, G., and A. Mantoglou (2013), Development of a multi-objective optimization algorithm using surrogate models for coastal aquifer management, Journal of Hydrology, 479, 13 – 23, doi: 10.1016/j.jhydrol.2012.10.050.

Koziel, S., and L. Leifsson (2012), Simulation-driven design using surrogate-based optimization and variable-resolution computational fluid dynamic models, Journal of Computational Methods in Science and Engineering, 12 (1-2), 75 – 98, doi:10.3233/JCM-2012-0405 .

Leary, S. J., A. Bhaskar, and A. J. Keane (2003), A knowledge-based approach to response surface modelling in multifidelity optimization, J. Global Optim., 26(3), 297–319.

Lilburne, L., and S. Tarantola (2009), Sensitivity analysis of spatial models, International Journal of Geographical Information Science, 23, 151 – 168.

Liong, S. Y., S. T.Khu, and W. T.Chan (2001), Derivation of Pareto front with genetic algorithm and neural network, J. Hydrol. Eng., 6(1), 52–61.

Mckay, M.D., R. J. Beckman, and W. J. Conover (1979), A comparison of three methods for selecting values of input variables from a computer code, Technometrics, 21(2), 239–245.

Madsen, J. I., and M. Langthjem (2001), Multifidelity response surface approximations for the optimum design of diffuser flows, Optim. Eng., 2, 453–468.

Matott, L. S., and A. J. Rabideau (2008), Calibration of complex subsurface reaction models using a surrogate-model approach, Adv. Water Resour., 31(12), 1697–1707, doi:10.1016/j.advwatres.2008.08.006.

Mayer, A. S., C. T. Kelley, and C. T. Miller (2002), Optimal design for problems involving flow and transport phenomena in saturated subsurface systems, Advances in Water Resources, 25, (8–12), 1233 – 1256.

Mondal, A., Y. Efendiev, B. Mallick, and A. Datta-Gupta (2010), Bayesian uncertainty quantification for flows in heterogeneous porous media using reversible jump Markov chain Monte Carlo methods, Adv. Water Resour., 33(3), 241–256.

Mugunthan, P., C. A.Shoemaker, and R. G.Regis (2005), Comparison of function approximation, heuristic, and derivative-based methods for automatic calibration of computationally expensive groundwater bioremediation models, Water Resour. Res., 41, W11427, doi:10.1029/2005WR004134.

Myers, R. H., and D. C. Montgomery (1995), Response surface methodology: process and product optimization using designed experiments, Wiley, New York

Pawar, R., G. Bromhal, S. Carroll, S. Chu, R. Dilmore, J. Gastelum, C. Oldenburg, P. Stauffer, Y. Zhang and G. Guthrie 2014. Quantification of Key Long-term Risks at CO2 Sequestration Sites: Latest Results from US DOE's National Risk Assessment Partnership (NRAP) Project. Energy Procedia 63: 4816-4823.

Qian, Z. G., C. C. Seepersad., V. R. Joseph, J. K. Allen, and J. C. F. Wu (2006), Building Surrogate Models Based on Detailed and Approximate Simulations, ASME Journal of Mechanical Design, 128, 668 – 677.

Qian, P. Z. G., and J. C. F. Wu (2008), Bayesian Hierachical Modeling for Integrating Low-Accuracy and High-Accuracy Experiments, Technometrics, 50, 192 – 204.

Queipo, N. V., R. T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, and P. K. Tucker (2005), Surrogate-Based Analysis and Optimization, Progress in Aerospace Sciences, 41(1), 1–28, doi:10.1016/j.paerosci.2005.02.001

Ratto, M., A. Castelletti, and A. Pagano (2012), Emulation techniques for the reduction and sensitivity analysis of complex environmental models, Environmental Modelling & Software, 34, 1-4, doi: 10.1016/j.envsoft.2011.11.003.

Razavi, S., B. A. Tolson, and D. H. Burn (2012a), Review of surrogate modeling in water resources, Water Resour. Res., 48, W07401, doi:10.1029/2011WR011527.

Razavi, S., B. A. Tolson, and D. H. Burn (2012b), Numerical assessment of metamodelling strategies in computationally intensive optimization, Environmental Modelling & Software, 34, 67-86.

Regis, R. G., and C. A.Shoemaker (2007), A stochastic radial basis function method for the global optimization of expensive functions, INFORMS J. Comput., 19(4), 497–509.

Regis, R. G., and C. A.Shoemaker (2009), Parallel stochastic global optimization using radial basis functions, INFORMS J. Comput., 21(3), 411–426.

Robinson, T., M. Eldred, K. Willcox, and R. Haimes (2006), Strategies for multifidelity optimization with variable dimensional hierarchical models, AIAA Paper 2006-1819, 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Newport, RI, 1–4 May 2006.

Sacks, J., W. J. Welch, T. J. Mitchell, and H. P. Wynn (1989), Design and analysis of computer experiments. Statist. Sci. 4 409-423.

Saltelli, A. (2002), Making best use of model valuations to compute sensitivity indices, Computer Physics Communications, 145, 280–297.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola (2008), Global Sensitivity Analysis. The Primer. John Wiley and Sons, Ltd, New York.

Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto (2004), Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models. John Wiley and Sons, Ltd, New York.

Saltelli, A., P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola (2010), Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, Computer Physics Communications, 181, 259–270.

Scollo, S., S. Tarantola, C. Bonadonna, M. Coltelli, and A. Saltelli (2008), Sensitivity analysis and uncertainty estimation for tephra dispersal models, J. Geophys. Res., 113, B06202, doi:10.1029/2006JB004864.

Schultz, M. T., M. J. Small, P. S. Fischbeck, and R. S. Farrow (2006), Evaluating response surface designs for uncertainty analysis and prescriptive applications of a large-scale water quality model, Environ. Model. Assess., 11(4), 345–359.

Schwarz, G. (1978), Estimating the dimension of a model, The Annals of Statistics, 6, 461–464.

Shrestha, D. L., N. Kayastha, and D. P. Solomatine (2009), A novel approach to parameter uncertainty analysis of hydrological models using neural networks, Hydrol. Earth Syst. Sci., 13(7), 1235–1248.

Simpson T. W., J. Peplinski, P.N. Koch, J. K. Allen (2001), Metamodels for computer-based engineering design: survey and recommendations, Eng. Comput., 17(2), 129–150.

Simpson, T. W., V. Toropov, V. Balabanov, F. A. C. Viana (2008), Design and analysis of computer experiments in multidisciplinary design optimization: a review of how far we have come – or not, *in*: Proceedings of the 12th AIAA/ISSMO multidisciplinary analysis and optimization conference, No. AIAA 2008–5802, Victoria, Canada, 2008.

Sobol, I. M., V. I. Turchaninov, Y. L. Levitan, B. V. Shukhman (1992), Quasirandom Sequence Generators, Keldysh Institute Of Applied Mathematics Russian Academy Of Sciences, Moscow (Russia), IPM ZAK. no.30.

Sobol, I. (1993), Sensitivity analysis for non-linear mathematical models, Mathematical Modelling and Computational Experiment 1, pp. 407–414. Translated from Russian: Sobol’,

I. M., 1990, Sensitivity estimates for nonlinear mathematical models, Matematicheskoe Modelirovanie 2, 112 – 118.

Sreekanth, J., and B.Datta (2011), Coupled simulation-optimization model for coastal aquifer management using genetic programming-based ensemble surrogate models and multiple-realization optimization, Water Resour. Res., 47, W04516, doi:10.1029/2010WR009683.

Sun, G., G. Li, M. Stone, and Q. Li (2010), A two-stage multi-fidelity optimization procedure for honeycomb-type cellular materials, Comput. Mater. Sci., 49(3), 500 – 511.

Sun, Y., C. Tong, Q. Duan, T. A. Buscheck, and J. A. Blink (2012), Combining Simulation and Emulation for Calibrating Sequentially Reactive Transport Systems. Transport in Porous Media, 92(2), 509 – 526.

Viana, F. A. C., V. Steffen, Jr., S. Butkewitsch, and M. D. F. Leal (2009), Optimization of aircraft structural components by using nature-inspired algorithms and multi-fidelity approximations, J. Global Optim., 45(3), 427 – 449.

Vitali, R., R. T. Haftka, and B. V. Sankar (2002), Multi-fidelity design of stiffened composite panel with a crack. Structural and Multidisciplinary Optimization, 23(5), 347 – 356.

Wang, G. G., and Shan S. (2007), Review of metamodeling techniques in support of engineering design optimization, J. Mech. Des., 129(1), 370 – 380.

Zhao, D., and D., Xue (2010), A comparative study of metamodeling methods considering sample quality merits, Structural and Multidisciplinary Optimization , 42(6), 923 – 938.

Tables

Table 1: Lithologies associated with the hydrostratigraphic units of the test case.

Hydrostratigraphic unit	**Lithologies**	**Mean length (horizontal direction)**	**Mean length (vertical direction)**	**Volumetric proportion**
Unit 1	Sand, coarse sand, medium sand, sand with gravel, gravel with sand, medium gravel, gravel, coarse gravel	717.7 m	8.3 m	0.60
Unit 2	Fine sand, very fine sand, silty sand, silt, silty clay, shale, sandstone, caliche, gypsum rock, clay, limestone.	478.5 m	5.6 m	0.40

Table 2: Input parameters for the LFM-2 base case run

Parameter	**Base case value**
Porosity (unit 1)	0.250
Porosity (unit 2)	0.330
Rock density (unit 1)	2400 kg/m^3
Rock density (unit 2)	2400 kg/m^3
Permeability (unit 1)	3.162x10^{-11} m^2
Permeability (unit 2)	3.162x10^{-17} m^2
Van Genuchten parameter m (unit 1)	0.655
Van Genuchten parameter m (unit 2)	0.190
Van Genuchten parameter alpha (unit 1)	5.62x10^{-5} m^{-1}
Van Genuchten parameter alpha (unit 2)	1.51x10^{-5} m^{-1}

Table 3: Input parameters ranges for the development of the linking functions for the test case

Parameter	Range (min – max)	Model
Porosity (unit 1)	0.25 – 0.50	HFM, LFM-2
Porosity (unit 2)	0.33 – 0.60	HFM, LFM-2
Rock density (unit 1)	2400 - 2800 kg/m^3	HFM, LFM-2
Rock density (unit 2)	2400 - 2800 kg/m^3	HFM, LFM-2
Permeability (unit 1)	-13.5 – -10.5* $log(m^2)$	HFM, LFM-2
Permeability (unit 2)	-15.0 – -18.0* $log(m^2)$	HFM, LFM-2
Van Genuchten parameter m (unit 1)	0.52 – 0.79	HFM, LFM-2
Van Genuchten parameter m (unit 2)	0.06 – 0.32	HFM, LFM-2
Van Genuchten parameter alpha (unit 1)	-4.69 – -3.81* $log(m^{-1})$	HFM, LFM-2
Van Genuchten parameter alpha (unit 2)	-5.50 – -4.14* $log(m^{-1})$	HFM, LFM-2
CO_2 leakage rate scaling parameter[1]	0.1 – 1.0	HFM, LFM-2
Brine leakage rate scaling parameter[2]	0.1 – 1.0	HFM, LFM-2
Chloride concentration in brine	-2.0 – 1.0* log(mol/L)	HFM, LFM-1
Arsenic concentration in brine	-9.0 – -5.0* log(mol/L)	HFM, LFM-1
Calcite initial volume fraction	0 – 0.2	HFM, LFM-1
Sorption scaling parameter[3]	-2.0 – 2.0*	HFM, LFM-1

*indicates log_{10} values.

[1]this factor was applied to the maximum CO_2 leakage rate

[2]this factor was applied to the maximum brine leakage rate

[3]this factor was applied to the adsorption capacity of different mineral phases.

List of Figures

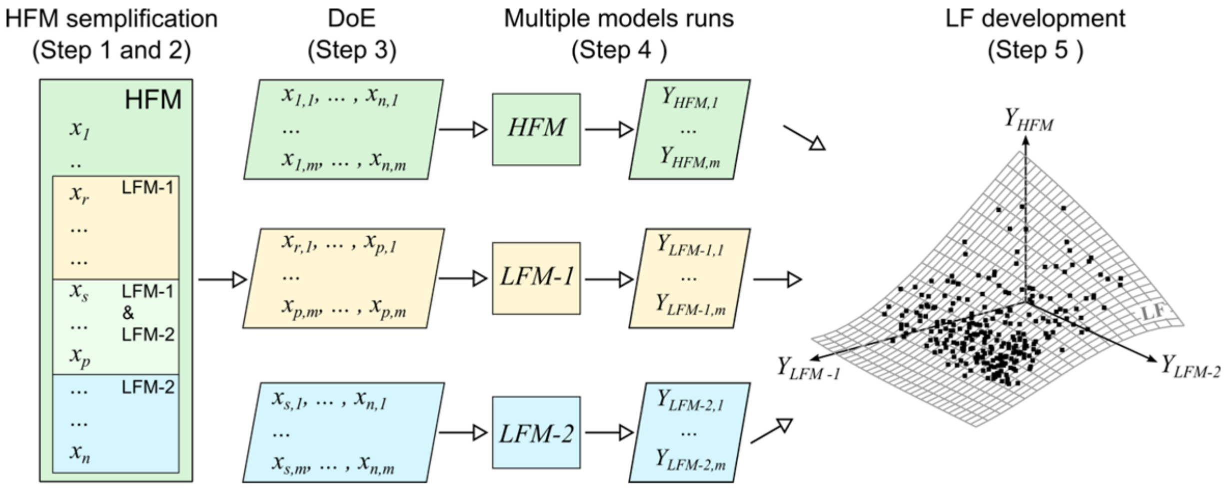

Figure 1. Linking function surrogate modeling (LFSM) framework.

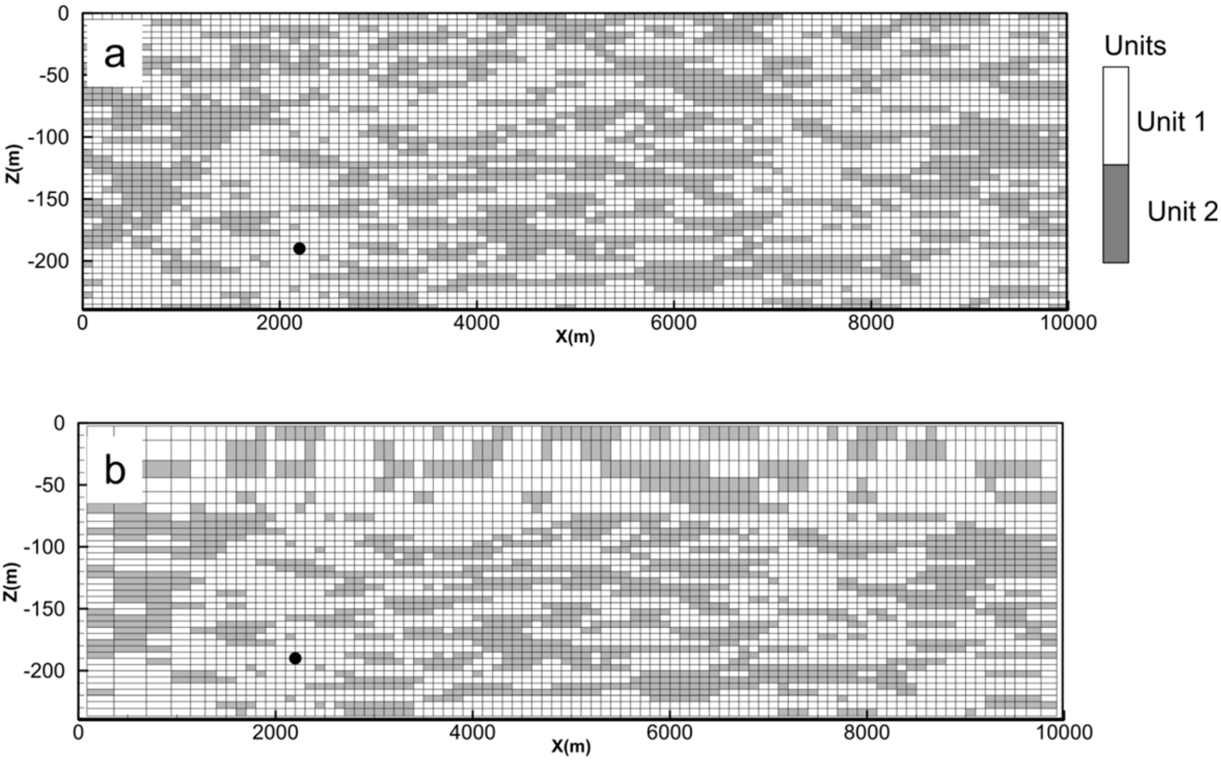

Figure 2: Heterogeneous distribution of two hydrostratigraphic units generated with T-PROGS (a). Numerical model mesh used for the TOUGHREACT simulations (b). The black circle indicates the location of the CO_2 and brine leakage point.

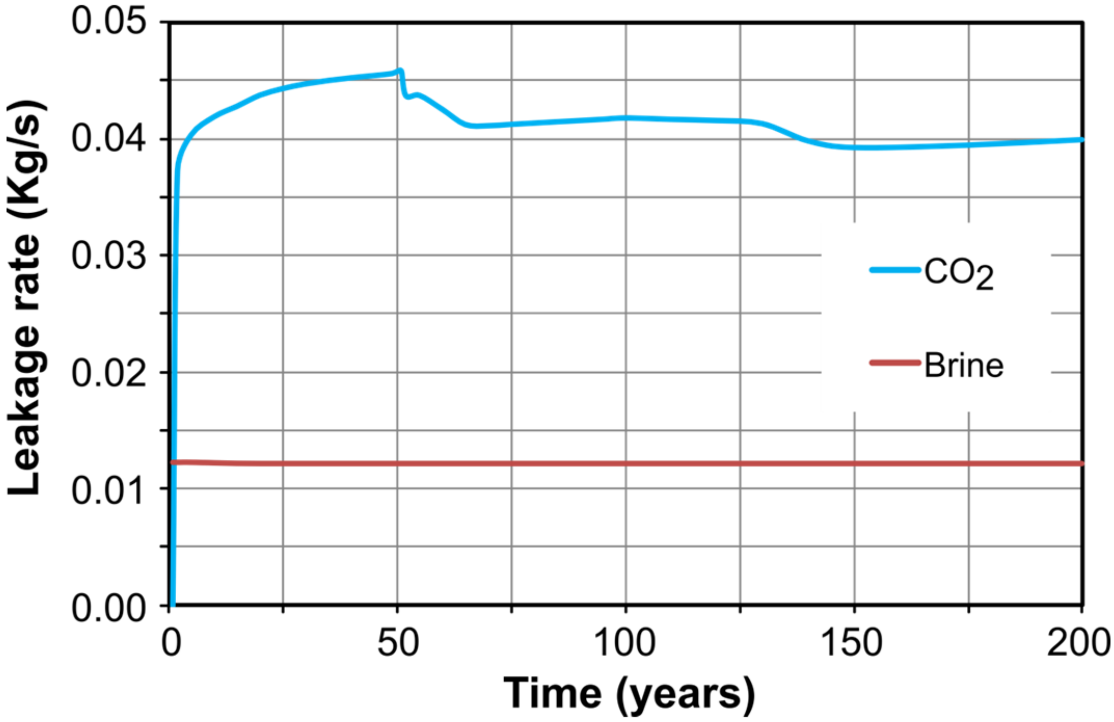

Figure 3: CO_2 and brine leakage rates over time.

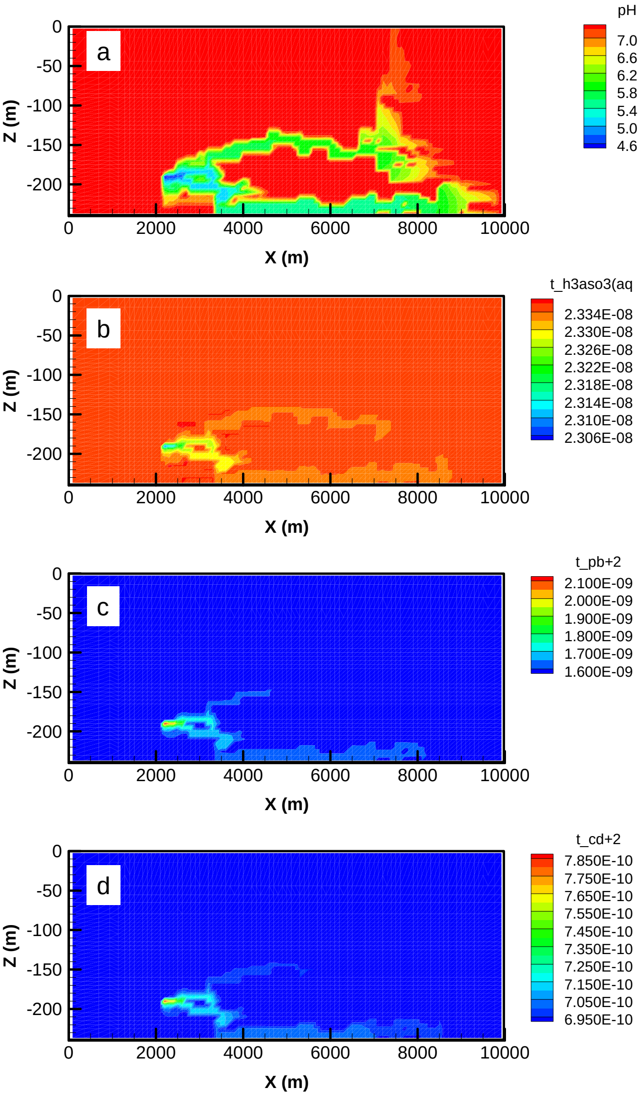

Figure 4: Results of the LFM-2 base-case simulation after 200 years of continuous leakage (unreactive transport). pH distribution (a), AsO_3 concentration (b), Pb^{2+} concentration (c), Cd^{2+} concentration (d).

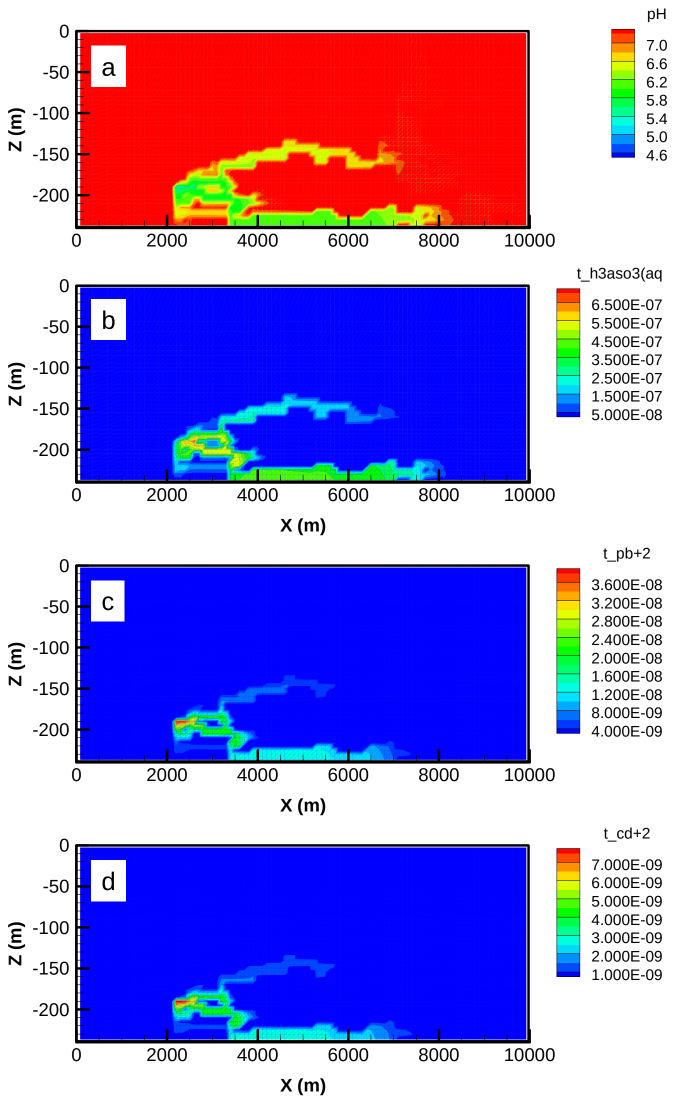

Figure 5: Results of the HFM base-case simulation after 200 years of continuous leakage (reactive transport). pH distribution (a), AsO_3 concentration (b), Pb^{2+} concentration (c), Cd^{2+} concentration (d).

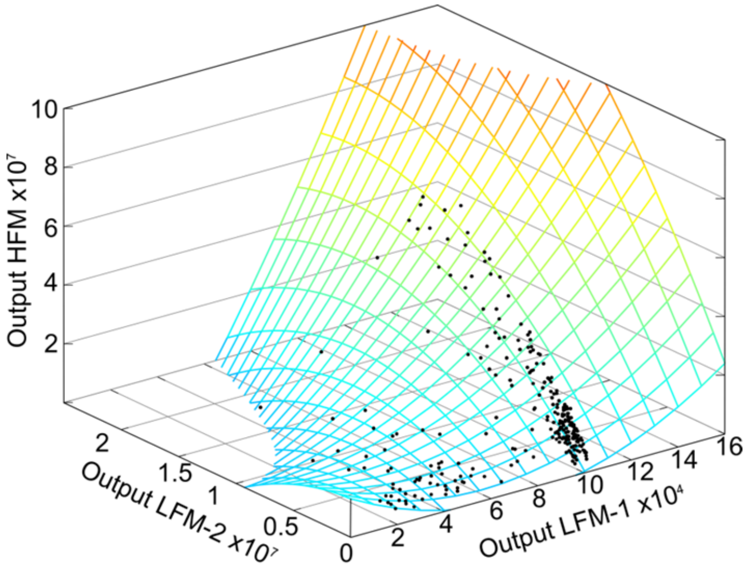

Figure 6. Second order polynomial linking function for estimating the volume of pH < 6.5 (m^3) after 180 days of leakage. Points represent the calculated responses from the two lower fidelity models (LFM-1 and LFM-2) and those from the high fidelity model (HFM).

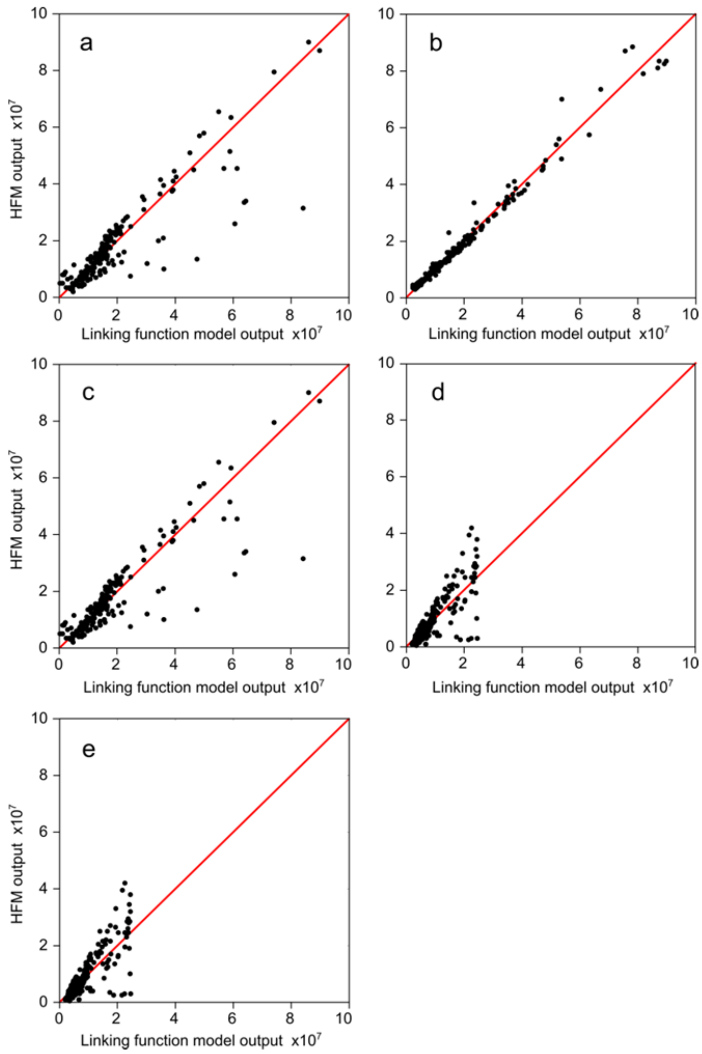

Figure 7: Comparison between HFM model outputs (volumes in m^3) and linking functions responses. Simulation time is 200 years. (a) pH (R^2 = 0.822); (b) TDS(R^2= 0.987); (c) As (R^2 = 0.723); Pb (R^2 = 0. 638); Cd (R^2 = 0. 635).

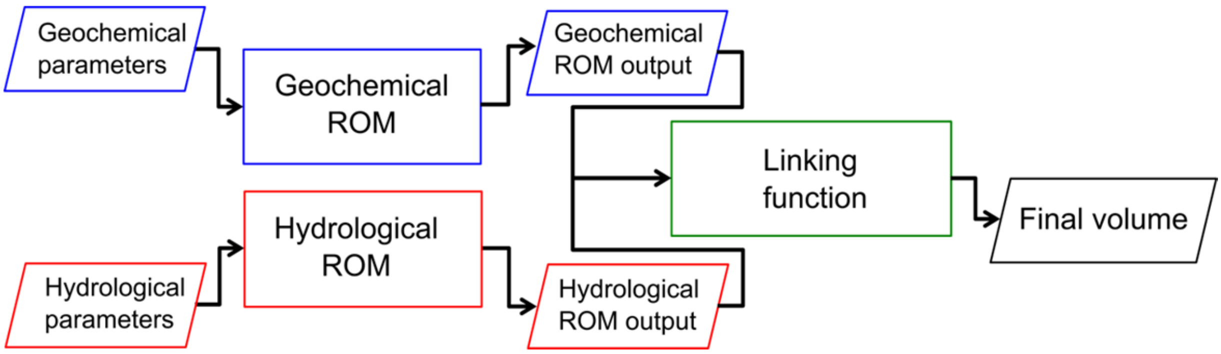

Figure 8: Flow chart for applying the linking function approach.