Dolomites Research Notes on Approximation
Proceedings of DWCAA12, Volume 6 · 2013 · Pages 140–150

Model calibration of a geomechanical problem with efficient global optimization

Alberto Lovison[a] · Francesco Comola[b] · Pietro Teatini[c] · Carlo Janna[c] · Massimiliano Ferronato[c] · Mario Putti[a] · Giuseppe Gambolati[c]

Abstract

Efficient Global Optimization (EGO) is an optimization strategy based on approximating functions, namely Gaussian process models. We show the application of this technique to a model calibration problem referred to a geomechanical application. By means of the approximating function an objective relevance ranking among the problem parameters can be produced, offering valuable and reusable information on the physical problem.

1 Introduction

The scope of this paper is to illustrate the effectiveness of a global optimization strategy based on approximating functions on a problem of geomechanical model calibration. Approximating functions, also referred to as *metamodels*, when they serve as a surrogate for a time consuming numerical model of a complex physical phenomenon, have proven to be very effective in diverse engineering problems, and their application has become wider and wider in recent years [1, 2, 3, 4, 5, 6, 7, 8, 9]. The method considered has proven to be very effective in global optimization problems, and in the proposed application is employed for solving an inverse problem, namely for determining unknown parameters of the numerical model representing important physical quantities, by matching the predictions of the numerical model with experimental measurements. As a byproduct, the method offers a detailed statistical analysis of the unknown function, which allows a ranking of the relative importance of the model parameters in terms of how much the function is affected by their variation.

As a possible drawback, the method requires a careful supervision from the experimenter, which must decide whether or not the approximations are sufficiently accurate for the application purposes. This problem potentially occurs for every situation involving physical quantities, and without some kind of global information (e.g., Lipschitz constant, bound on second derivative, functional form, et cetera), which usually is not available, one is never guaranteed to have detected the global optimum of an unknown function, unless the optimization process have explored exhaustively the search domain, i.e., unless the search sequence becomes everywhere dense [10].

2 Approximating unknown functions with Gaussian processes

There are several types of functions which can be effectively used for approximating general unknown functions, e.g., neural networks, splines, Fourier polynomials, radial basis functions, and many others. Nevertheless, it can be proved that neural networks, radial basis functions and gaussian processes are equivalent, in the sense that approximations obtained with each of these methods are equally accurate. We have chosen Gaussian processes, because there is a vast literature discussing their application and performances, efficient numerical implementations are available and finally because they have the special feature of offering both predictions and error estimates. Incidentally, Gaussian processes are universally used in geostatistics, where they are referred to as the *kriging* method, because of the work of Prof. Daniel Krige.

Formally, [11, Chapter 2, page 13]

DEFINITION 1. *A* Gaussian process *is a collection of random variables, any finite number of which have a joint Gaussian distribution.*

A Gaussian process is completely specified by its mean and covariance functions

$$m(\mathbf{x}) = \mathbb{E}(f(\mathbf{x})), \tag{1}$$

$$\operatorname{cov}(\mathbf{x},\mathbf{x}) = \mathbb{E}\left((f(\mathbf{x}) - m(\mathbf{x}))\cdot(f(\mathbf{x}') - m(\mathbf{x}'))\right), \qquad \mathbf{x},\mathbf{x}' \in \mathbb{R}^d. \tag{2}$$

The type of covariance function determines the smoothness of the Gaussian process. One of the most used classes of covariance functions is the *squared exponential*:

$$\operatorname{cov}_{SE}(\mathbf{x},\mathbf{x}') := \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\ell^2}\right), \tag{3}$$

[a] M3E S.r.l., Padova, Italy and Department of Mathematics, University of Padova, Padova, Italy.
[b] M3E S.r.l., Padova, Italy and CRYos Lab., Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland.
[c] M3E S.r.l., Padova, Italy and Department of Civil, Environmental and Architectural Engineering, University of Padova, Padova, Italy.

which induces the process to be infinitely differentiable in mean square. Mean and covariance determine the smoothness and the error magnitude of the predictions that will be produced by a Gaussian process. A random field is defined also before any function evaluation being performed. Such random field is also called the prior distribution, or simply the *prior*.

When the unknown function has been evaluated at a number of sites, the so called training sample of points or starting design, one can apply Bayes' rule and obtain the *posterior distribution*, for any unknown site, which is the probability distribution given by the prior conditioned on the information provided in the starting design.

The expected value of the posterior distribution at any untried site is used as a prediction for the value of the unknown function, while the standard deviation is used as an error estimate. The determination of the type of covariance function and its parameters (also called meta–parameters) on the basis of the function evaluations on the starting dataset is also another interesting aspect of the problem of approximation and optimization [5, 4, 11, 12].

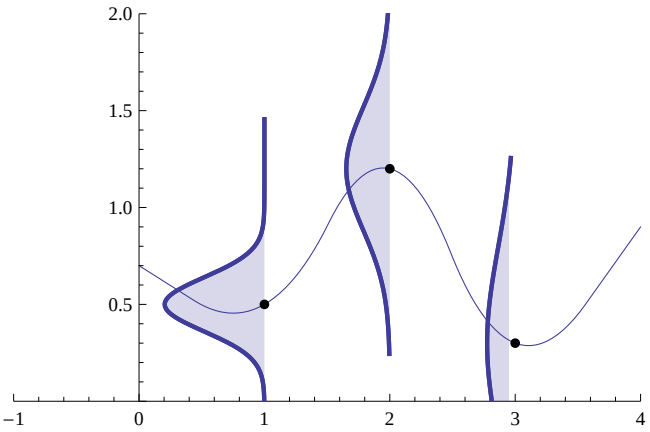

Figure 1: Example of Gaussian process: at every point in the domain corresponds a predictive probability distribution, from which it is possible to extract mean (function prediction) and standard deviation (error estimate)

A Gaussian process metamodel can be implemented efficiently as follows. Let $X = \left\{\mathbf{x}_i = (x_1^i, \ldots, x_d^i)^\top \in \mathbb{R}^d : i = 1, \ldots, n\right\}$ the starting sample, i.e., the set of d–tuples of coordinates corresponding to the points in the d–dimensional domain chosen for function evaluation. Let also $\mathbf{x}^\star$ be the coordinate vector of a generic untried point in the domain. Let $y = (y_1, \ldots, y_n)^\top$ the n–tuple of the function values corresponding to the points in the starting sample:

$$y_i := f(\mathbf{x}_i) = f(x_1^i, \ldots, x_d^i), \qquad i = 1, \ldots, n. \tag{4}$$

We set K be the simmetric positive definite matrix of the covariances[1] between points in the starting sample,

$$K := \left(\operatorname{cov}(\mathbf{x}_i, \mathbf{x}_j)\right)_{i,j=1,\ldots,n}, \tag{5}$$

while the vector $\mathbf{k}^\star := (\operatorname{cov}(\mathbf{x}_1, \mathbf{x}^\star), \ldots, \operatorname{cov}(\mathbf{x}_n, \mathbf{x}^\star))^\top$.

The predictive distribution of the function value $f^\star$ at the point $\mathbf{x}^\star$ given the observations $\mathbf{y}$ at the points X is given by:

$$f^\star \,\Big|\, X, \mathbf{y}, \mathbf{x}^\star \sim \mathcal{N}(\bar{f}^\star, \operatorname{cov}(f^\star)), \quad \text{where,} \tag{6}$$

$$\bar{f}^\star := \mathbb{E}\left[f^\star \,\middle|\, X, \mathbf{y}, \mathbf{x}^\star\right] = \mathbf{k}^{\star\top} K^{-1} \mathbf{y}, \tag{7}$$

$$\mathbb{V}[f^\star] := \operatorname{Var}\left(f^\star \,\middle|\, X, \mathbf{y}, \mathbf{x}^\star\right) = \operatorname{cov}(\mathbf{x}^\star, \mathbf{x}^\star) - \mathbf{k}^{\star\top} K^{-1} \mathbf{k}^\star. \tag{8}$$

The numerical implementation of this algorithm is based on the Cholesky decomposition, which can be performed very efficiently and is considered very stable.[2]

3 Efficent global optimization

3.1 Global Exploration not only local exploitation

Global optimization deals with the problem of organizing an efficient search strategy for finding the global optimum for a possibly multimodal unknown function over a compact domain. The Newton method is known as a very efficient method for finding a local optimum, likely the optimum in the basin of attraction containing the starting point. The optimum reached depends on the starting point, and therefore the Newton method is classified as a *local exploitation* strategy (see Figure 2).

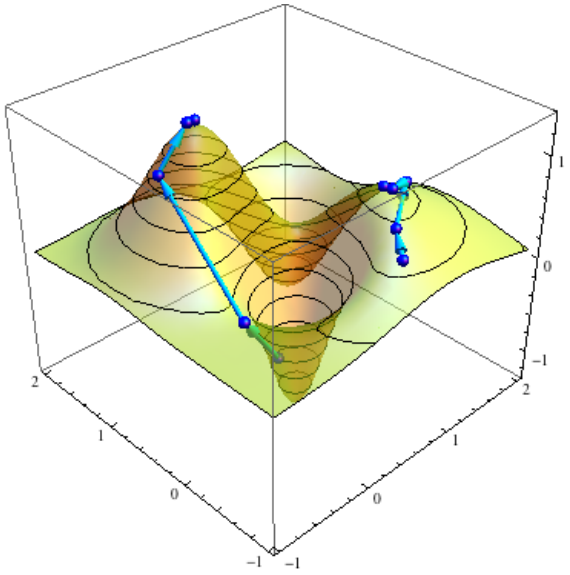

Figure 2: Gradient based optimization strategies, like the Newton method, approach fast to local optima, but the optimum reached is not guaranteed to be the global optimum, because it depends on the starting point.

On the other hand, if a search method is defined to become everywhere dense in infinite time, we would be guaranteed that the global optimum of the unknown function would be sooner or later approximated with arbitrary precision. Any such strategy would be classified as a *global exploration* strategy. Defining efficient global optimization strategies is a fascinating problem, and there is a wide literature discussing several methods involving approximating functions, in particular Gaussian processes.[3]

3.2 Global optimization with Gaussian processes

Maybe the simplest way of using approximations for optimization consists in successively add as a new point the optimum of the approximate function fitted on the evaluations available so far. An almost exhaustive search on the problem domain should be affordable because function evaluations of the metamodel are inexpensive. This strategy could be very efficient on the point of view of local exploitation, while it would not guarantee to detect the global optimum. Indeed unexplored regions with sub optimal predictions will not likely be selected by such an algorithm, while they could conceive marked variations and even the global optimum.

A very interesting global optimization algorithm called EGO (Efficient Global Optimization) involves both the prediction and the error estimate in order to become everywhere dense in infinite time, while maintaining the desirable feature of rapidly approaching local optima.

3.3 Expected improvement

EGO is an algorithm that iteratively updates an approximation of the unknown function by adding a new point in the maximum of an auxiliary function called *expected improvement*. The expected improvement is a real number measuring the probability that the value of the unknown function at an untried site will be better than the optimum obtained so far in the search process (see Figure 3).

More precisely, assume we are trying to *minimize* the function f, and let $f_1, \dots, f_n$ the function values computed at the n-th step of the search process, i.e., the values obtained computing f on the first n sites $\mathbf{x}_1, \dots, \mathbf{x}_n$. Let $min_{f,n} := \min_{i=1,\dots,n} f_i$ the minimum value obtained so far. Let us fix a generic untried site $\mathbf{x}$ and let $F := f \,\big|\, X, \mathbf{y}, \mathbf{x}$ be the posterior distribution at a site $\mathbf{x}$ as defined in (6–8). The *improvement* $I(\mathbf{x})$ at $\mathbf{x}$ is a random variable given as

$$I(\mathbf{x}) := \max\left\{min_{f,n} - F, 0\right\}, \tag{9}$$

thus the expected improvement is

$$\mathbb{E}[I(\mathbf{x})] := \mathbb{E}[\max\left\{min_{f,n} - F, 0\right\}]. \tag{10}$$

[1] The positivity is a consequence of the functional form of the covariance in (3). Thanks to the positivity the covariance function is consistent with (2) automatically. Further valid functional forms for the covariant functions are described in Chapter 4 of [11].

[2] A number of free and commercial implementations of Gaussian processes are available on the web, along with detailed documentation, e.g., PErK[12], gpml[11], modeFRONTIER® and the DACE toolbox for MATLAB® and SCILAB®.

[3] See, e.g., [4] and the references therein.

We can obtain a closed form for the expected improvement by recalling that F is a normally distributed random variable, and by writing m_F for its mean value and σ_F for its standard deviation. If we denote with ϕ the probability density function of the standard normal distribution, and with Φ its cumulative distribution function, we have that

$$\mathbb{E}[I(\mathbf{x})] = (min_{f,n} - m_F)\Phi\left(\frac{min_{f,n} - m_F}{\sigma_F}\right) + \sigma_F \phi\left(\frac{min_{f,n} - m_F}{\sigma_F}\right). \tag{11}$$

For a detailed discussion of this method the reader is referred to [5].

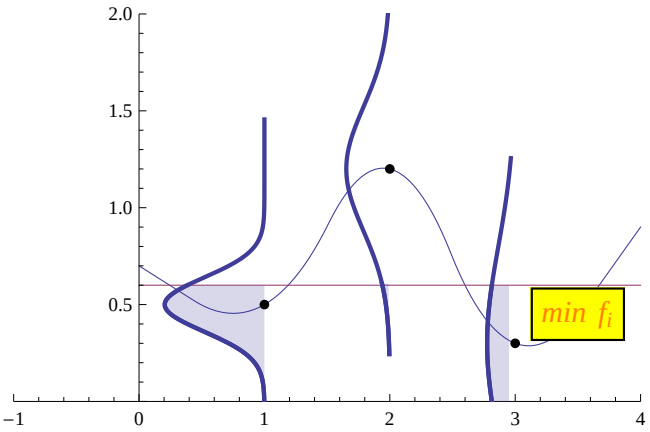

Figure 3: Graphical representation of the expected improvement for the Gaussian process example of the previous section. The thick blue curves displayed vertically represent the p.d.f. of the posterior distributions given by the Gaussian process at different sites. The purple horizontal line represents the minimum of the function obtained so far and the expected improvement at each site is represented by the blue area.

A straightforward computation of the derivatives in (11) shows that the expected improvement is monotonically decreasing with the prediction m_F and monotonically increasing with the standard deviation σ_F. Under some non degeneracy conditions, it is possible to prove that unexplored regions, which have large σ_F, will be sooner or later sampled by maximizing the expected improvement, and, therefore, the domain will be sampled everywhere densely in infinite time and global convergence will be guaranteed.[4]

4 Efficient Global Optimization in Action

As an application we illustrate a model calibration problem for an expensive numerical simulation of the behavior of an underground gas reservoir when large volumes of gas are periodically injected and extracted [15, 16]. The model calibration is performed by fitting millimeter precise radar satellite measurements of vertical and horizontal ground displacements. Maps of the ground vertical and horizontal displacements detected by satellite measurements (PSInSAR) used in the present application are reported in Figure 4.

The geometry of the geomechanical model is based on a three dimensional seismic survey of the underground. The unknowns of the model are the transversally isotropic elastic parameters, and are determined by fitting the predictions with the satellite measurements, by using the known injection pressure as driving force. One of the main novelties of the proposed geomechanical model is its anisotropy, which causes us to deal with five model parameters, only one of which can be fixed according to physical considerations. We remain with a four dimensional parameters space along with a nonlinear constraint.

In similar situations, engineers and experimenters can determine one or two parameters correctly by using their experience, the knowledge of the physical problem and standard trial and error procedures. More parameters, especially when multimodality potentially occurs, are harder to fit and results are less reliable. Furthermore, such handcrafted optimization, though possibly very effective, could not offer reusable information for similar problems which likely may occur in future.

Global optimization can offer a principled methodology for calibrating parameters. Furthermore, by using the metamodel fitted during the optimization process, one can perform a detailed statistical analysis of the importance of the parameters on the objective function, giving a ranking among them and an estimate of their reciprocal correlation (see Section 7).

[4] Actually, it suffices to suppose that the function f is bounded and continuos, and moreover that the characteristic length ℓ in (3) of the gaussian process model is not zero. In such a situation the sampling sequence will be everywhere dense in infinite time. (see [13, 14, 4]).

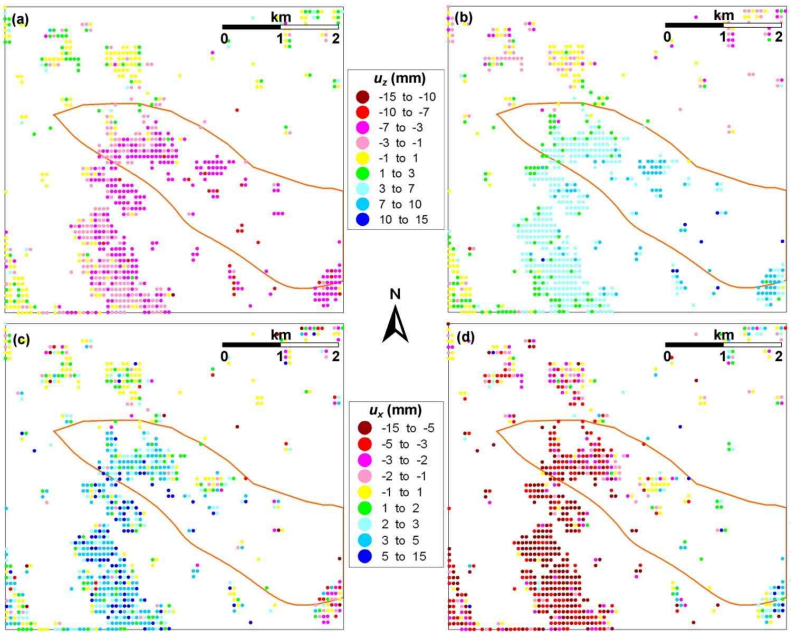

Figure 4: PSInSAR analysis of data acquired from november 2005 to april 2006 (left panels) and from april 2006 to november 2007 (right panels). The measured seasonal vertical displacements (upper panels) are of 8-10 mm and the horizontal displacements are of and 6-8 mm (lower panels). The trace of the gas reservoir is shown by the orange line.

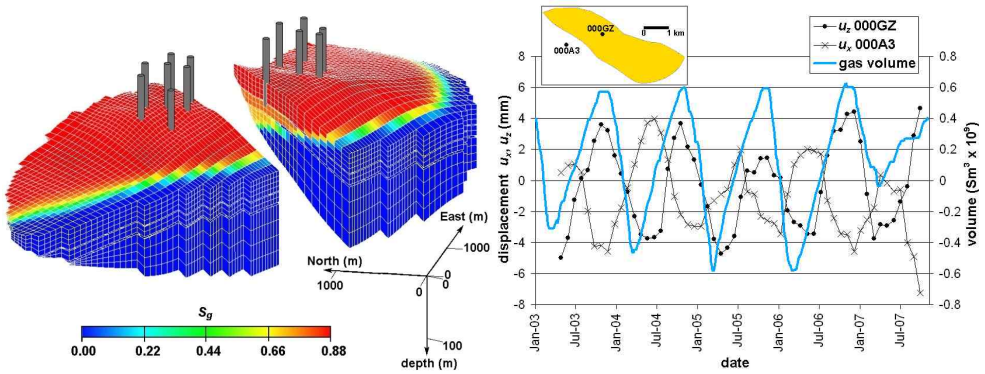

Figure 5: Panel (a): numerical model of the gas/water interaction. Panel (b): SAR measurement of the seasonal ground displacements due to injection–extraction of gas.

4.1 EGO in the geomechanical application

The EGO algorithm starts with a space–filling dataset, usually a latin hypercube design (LHD) with optimal maximin distance , which should offer a rough but rather complete overview of the problem domain.[5]

However, in the considered geomechanical problem, because of the presence of a nonlinear constraint, it was not possible to adopt a LHD, because the desired properties of marginal distributions are not preserved under constraining. Therefore, in order to define a space–filling design we have considered a regular grid (five nodes for each of the four parameters), we have extracted the subset fulfilling the constraint and taken the resulting design as the starting dataset (see Figure 7). A problem specific performance function has been defined in order to measure the matching between the model and the satellite measurements during injection-extraction cycles (see Figure 6).

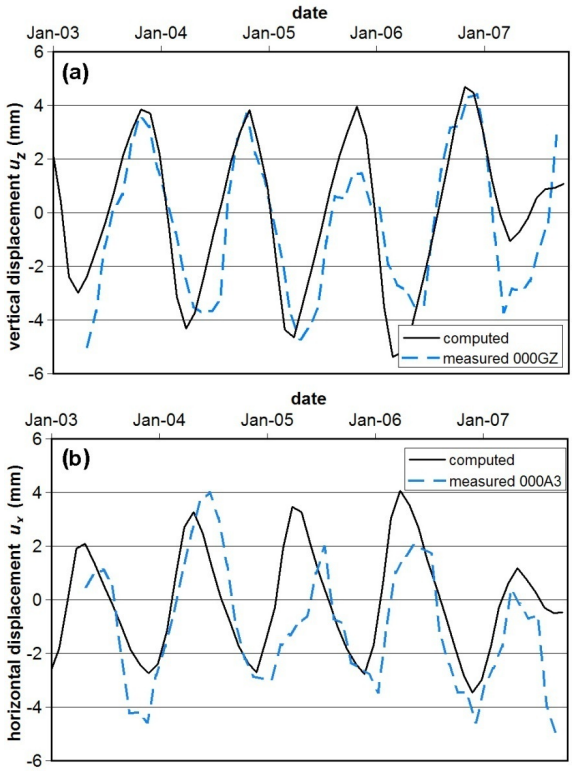

Figure 6: Matching between the ground displacement predicted by the geomechanical model and radar satellite measurements.

Typically, such performance functions are based on least squares, therefore the function values are positive and the optimum is obviously near zero. This is problematic when a metamodel is used for approximating and optimizing such a performance function, because it is clear that achieving small residuals in the vicinity of small values of the performance is far more important than obtaining residuals of the same magnitude for values far from zero. It is therefore recommendable to magnify the performance function values close to zero, by composition with a logarithmic transformation, or in alternative passing to reciprocals $y \mapsto -1/y$.

More technically, in order to establish if a function transformation is necessary, we perform a leave–one–out cross validation on the starting dataset and verify if the cross validated residuals are distributed normally and are compatible with the Gaussian process predictions of the error.

5 Metamodel validation

After computing the performance function on the starting dataset we have fitted the function values with a Gaussian process model with independent scaling parameters (the parameter ℓ in (3)) for every variable, i.e., an anisotropic model. Then a metamodel validation has been performed by analyzing the distribution of the *cross-validated residuals*, i.e., the discrepancy from the true value of the function at a dataset point and the prediction of a metamodel fitted to the remaining

[5] A latin hypercube design is a very effective method for producing space filling samples in an n-dimensional hyper-rectangle: it has good marginal properties, i.e., the projection of an m points design in any single variable is always a sequence of m equally spaced points. Furthermore, generating an LHD of any size in any dimension is almost computationally inexpensive. There are many of these designs, however these could be clusterized in space or leave important lacks in the domain. If the design is generated taking into account some reciprocal distance criterion, also those problems are avoided. It is a common practice to choose a maximin LHD as a starting sample for fitting a gaussian process to an unknown function. See [17, 18, 19, 3, 12].

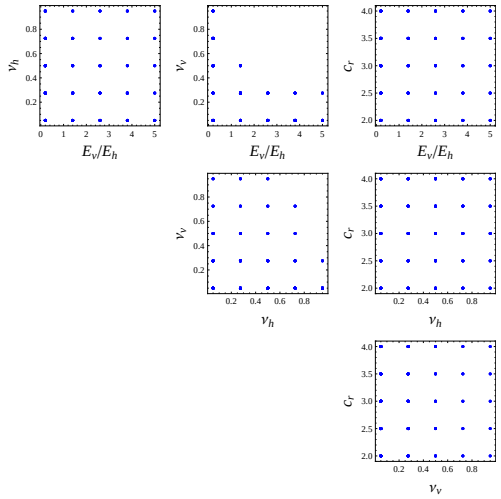

Figure 7: Orthogonal projections on pairs of coordinates of the starting dataset of the geomechanical problem. The dataset is composed by 265 points obtained imposing the nonlinear constraint to a regular grid with five nodes in each variable ($5^4 = 625$).

N-1 points. These residuals are standardized by dividing by the standard deviation predicted by the metamodel σ_F. The standardized cross validated residuals should lie in a (-3, 3) interval, while in the quantile–quantile plot the data points should display along the principal bisector line (the matching would be perfect if the residuals were distributed normally). If those tests apparently fail it is possible that a suitable function transformation would fix the problem. Because for the typical error functions the most important values are close to zero, it is convenient to magnify small function values by a log transformation or alternatively a $-1/y$ transformation, as mentioned above.

In Figure 8 we show the results of the metamodel validation procedure on the starting dataset by using a reciprocal transformation ($y \mapsto -1/y$) of the performance function based on least squares. The reciprocal transformation is the choice that produced the best results, if compared with the untransformed function and with the logarithmic transformation. However, as it can be seen in the central panel, still some of the standardized residuals remain outside of the $(-3, 3)$ interval, while the tails of the distribution deviate from the diagonal of the quantile–quantile plot, suggesting that possibly the dataset used could be still too coarse or that it could happen that using Gaussian processes with less smoothness requirement could improve the metamodel fitting.

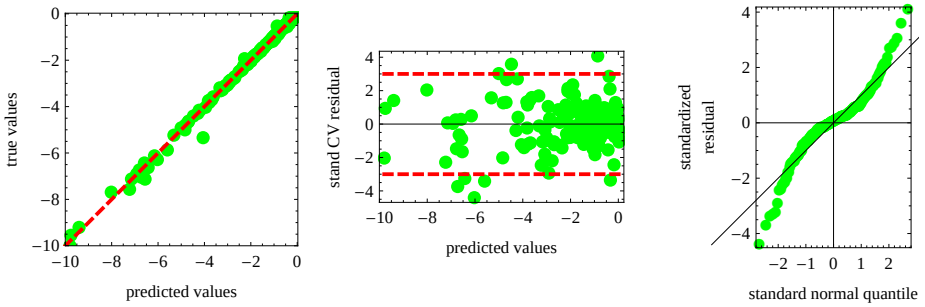

Figure 8: Metamodel validation for the $-1/y$ transformed merit function. Left: true values versus predicted values. Center: standardized residuals. Right: quantile–quantile plot of the standardized residuals.

6 Optimization

According to the EGO method, at each iteration step we should find the global maximum of the expected improvement on the basis of the Gaussian process model, incorporate this point to the dataset, fit a new model with the augmented dataset and restart by finding the global maximum of the expected improvement on the new model. Due to the limited time available for the entire process and in order to exploit the available parallel resources, we choose to limit ourselves to two iterations, and to compute a batch of candidate points, instead of computing a point at a time. Candidate points were found by performing a gradient based multistart search starting from the best performing points picked from a dense grid evaluated on the Gaussian process metamodel.

In Figure 9 (upper panel) the candidate points for the first iteration (10 points) and the second iteration (5 points) are shown.

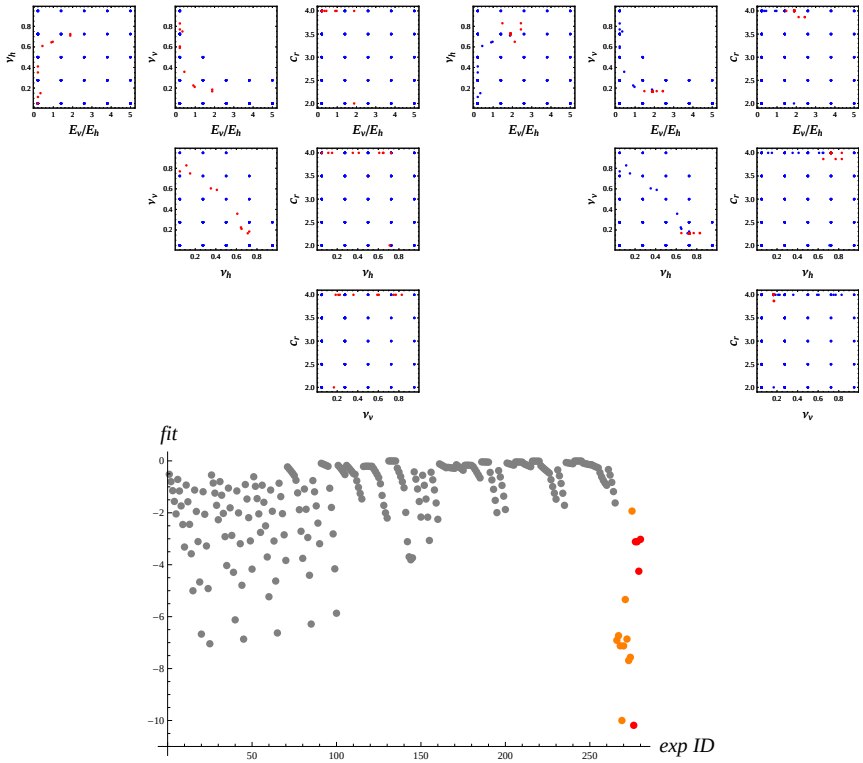

Figure 9: Upper panel: (left) first EGO iteration (10 points), (right) second EGO iteration (5 points). Lower panel: performance function values for the starting dataset (gray), the first iteration (orange) and last iteration (red).

The performance function values corresponding to starting dataset and subsequent iterations are reported in Figure 9 (lower panel), showing a marked improvement at every step, although obviously not all of the new points will perform better than the points of the previous batch, because also many points with large predicted uncertainty must be included for guaranteeing the detection of the global optimum. An additional set of computations has been reserved for the statistical analysis of the parameters importance.

7 Sensitivity Analysis with Functional ANOVA

As observed above, the nonlinear constraint on the domain prevented the use of a latin hypercube design, a space–filling design with good marginal properties also suitable for statistical analysis of the functional relation between parameters and outcome of the unknown function. We have therefore reserved a number of function evaluations for running an LHD surrounding the optimum obtained with the optimization, sufficiently small for being contained in the feasible region.

The statistical analysis of the parameters influence on the outcome is performed via a *functional ANalysis Of VAriance (ANOVA):* the total variance of the function $\mathbf{x} \mapsto y(\mathbf{x})$, where here $y(\mathbf{x})$ represents the outcome of the metamodel, is decomposed into contributions which can be ascribed to single variables or to couples of variables, and so on, in a way that is reminiscent of a Taylor expansion.

$$\mathbf{x} = (x_1, \ldots, x_d) \longmapsto y(\mathbf{x}), \tag{12}$$

$$y(\mathbf{x}) = y_0 + \underbrace{\sum_{i=1}^{d} y_i(x_i)}_{\text{main effects}} + \underbrace{\sum_{1 \leqslant i < j \leqslant d} y_{i,j}(x_i, x_j)}_{\text{interactions}} + \underbrace{\cdots + y_{1,\ldots,d}(x_1, \ldots, x_d)}_{\text{higher order terms}}. \tag{13}$$

The analysis of variance allows to rank the variables $x_1, \ldots, x_d$ by quantifying with an objective criterion their contribution to the total variance of the function over the domain. If recommendable, one could remove (i.e., fix to a nominal value) the less important variables for further analysis or exploration.

More precisely, we have fitted a new Gaussian process model on the LHD around the optimum, then we have estimated the contribution to the outcome of the ith parameter as follows. We define a sequence of equally spaced values for the ith

parameter in its own range, and for every fixed value we numerically integrate the model over the remaining parameters. If we consider for simplicity the unit d–dimensional hypercube $[0,1]^d$ as a domain, we can write:

$$y_i(x_i) = \int y(x_1,\dots,x_d)d\mathbf{x}_{-i} - y_0$$

$$\text{where} \quad y_0 := \langle y \rangle = \int_{[0,1]^d} y(\mathbf{x})d\mathbf{x},$$

$$\int y(x_1,\dots,x_n)d\mathbf{x}_{-i} := \int_{[0,1]^{d-1}} y(x_1,\dots,x_i,\dots,x_d)dx_1\cdots dx_{i-1}dx_{i+1}\cdots dx_d.$$

These averaged behaviors associated to each variable are called the *main effects* and for our example are plotted in Figure 10. Analogously, we can compute the contributions of higher order in the expansion (13), by fixing the values of a subset of

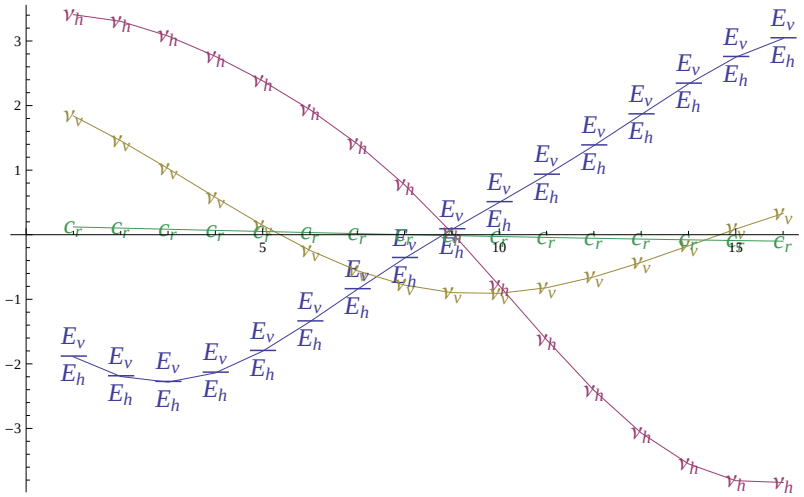

Figure 10: Functional ANOVA: main effects. Each curve corresponds to the effect of a single parameter on the unknown function, and is obtained averaging the metamodel over the remaining parameters in turn.

s parameters $x_{i_1},\dots,x_{i_s}$, averaging over the remaining, and subtracting the lower order effects. For instance, the second order terms are called *interactions*:

$$y_{i,j}(x_i,x_j) := \int y(x_1,\dots,x_d)d\mathbf{x}_{-(ij)} - y_0 - y_i(x_i) - y_j(x_j), \tag{14}$$

where $d\mathbf{x}_{-(ij)}$ denotes integration over all variables except x_i and x_j. Interactions for the geomechanical example are plotted in Figure 11.

The visualization of the main effects and interactions gives valuable and objective information on the problem at hand. Furthermore, one can estimate the absolute and relative importance of the parameters by computing total and partial variances for $y(\mathbf{x})$. The total variance is

$$V := \int_{[0,1]^d} y^2(\mathbf{x})d\mathbf{x} - y_0^2, \tag{15}$$

while partial variances are referred to the single terms in the expansion (13)

$$V_i := \int_0^1 y_i^2(x_i)dx_i, \tag{16}$$

$$V_{i,j} := \int_0^1\int_0^1 y_{i,j}^2(x_i,x_j)dx_i dx_j, \tag{17}$$

$$V_{i_1,\dots,i_s} := \int_0^1\cdots\int_0^1 y_{i_1,\dots,i_s}^2(x_{i_1},\dots,x_{i_s})d\mathbf{x}_{-(i_1,\dots,i_s)}. \tag{18}$$

By squaring and integrating both sides of (13) one obtains the formula for the variance decomposition,

$$V = \sum V_i + \sum V_{i,j} + \cdots + V_{1,\dots,d}, \tag{19}$$

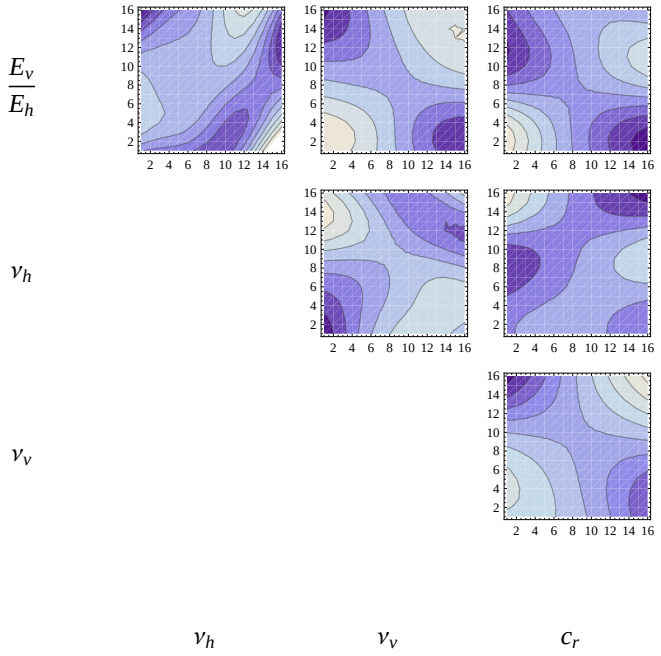

Figure 11: Functional ANOVA: interactions.

which legitimates the definition of the normalized variances, i.e., the *sensitivity indexes* $S_{i_1,\ldots,i_s}$:

$$S_{i_1,\ldots,i_s} := \frac{V_{i_1,\ldots,i_s}}{V}. \tag{20}$$

Because sensitivity indexes sum to 1, we have that S_i represent the proportion of the variation of y which is due to x_i, while $S_{i,j}$ represents the interaction between parameter x_i and parameter x_j which cannot be explained by their main effects, and so on. A summary of the analysis of variance (ANOVA) performed on the function example is reported in Figure 12. From the ANOVA table we deduce that the most important contributions to the variance are the main effects of ν_h and E_v/E_h and the interaction between E_v/E_h and ν_v. Furthermore, we see that to the pair $(\nu_h, E_v/E_h)$ can be ascribed the about the 75% of the effects on the outcome y.

Variable	S_i	Variable Pair	S_{ij}
$\frac{E_v}{E_h}$	22.5949%	$(\frac{E_v}{E_h}, \nu_h)$	5.02533%
ν_h	47.8916%	$(\frac{E_v}{E_h}, \nu_v)$	13.5289%
ν_v	4.68293%	$(\frac{E_v}{E_h}, c_r)$	0.045289%
c_r	0.0323321%	(ν_h, ν_v)	3.20165%
		(ν_h, c_r)	0.0604631%
		(ν_v, c_r)	0.0392705%

Figure 12: Functional ANOVA: summary of sensitivity indexes

8 Conclusions

An accurate calibration of the finite element geomechanical model to the SAR interferometry data has been performed by using an efficient global optimization strategy, called EGO, which proved to be valuable also in the presence of uncertain measurements. As a result, EGO makes possible to estimate the functional dependence of physically relevant quantities,

e.g., the maximum vertical seasonal displacement, with respect to operational parameters (e.g., working pressure) and to uncertain model parameters (e.g., elastic moduli), leading to predictions of the reservoir behavior under untested configurations. Furthermore, a statistical analysis of the functional dependence, called ANOVA, gives relevant and reusable information about the physical problem, which could be useful for designing more refined modelizations.

References

[1] J Sacks, W J Welch, T J Mitchell, and H P Wynn. Design and Analysis of Computer Experiments. *Statistical Science*, 4(4):409–423, 1989.

[2] V Picheny, D Ginsbourger, O Roustant, R T Haftka, and N-H Kim. Adaptive Designs of Experiments for Accurate Approximation of a Target Region. *Journal Of Mechanical Design*, 132(7):071008, 2010.

[3] A Lovison and E Rigoni. Adaptive sampling with a Lipschitz criterion for accurate metamodeling. *Communications in Applied and Industrial Mathematics*, 1(2):110–126, 2010.

[4] D R Jones. A taxonomy of global optimization methods based on response surfaces. *Journal of Global Optimization*, 21(4):345–383, 2001.

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

[6] M D Buhmann. *Radial basis functions: theory and implementations*, volume 12 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2003.

[7] H M Gutmann. A radial basis function method for global optimization. *DAMTP*, pages 1–29, November 2007.

[8] A Iske. *Multiresolution methods in scattered data modelling*, volume 37 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, Berlin, 2004.

[9] H Wendland. *Scattered data approximation*, volume 17 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2005.

[10] C P Stephens and W Baritompa. Global Optimization Requires Global Information. *Journal of Optimization Theory and Applications*, 96:575–588, 1998.

[11] C E Rasmussen and C K I Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2006.

[12] T J Santner, B J Williams, and W I Notz. *The design and analysis of computer experiments*. Springer Series in Statistics. Springer-Verlag, New York, 2003.

[13] M Locatelli. Bayesian Algorithms for One-Dimensional Global Optimization - Springer. *Journal of Global Optimization*, 1997.

[14] J P C Kleijnen, W Beers, and I Nieuwenhuyse. Expected improvement in efficient global optimization through bootstrapped kriging. *Journal of Global Optimization*, 54(1):59–73, 2012.

[15] G Gambolati, M Ferronato, and P Teatini. Reservoir compaction and land subsidence. *Revue Européenne de Génie Civil*, 10(6-7):731–762, 2006.

[16] P Teatini, N Castelletto, M Ferronato, G Gambolati, C Janna, E Cairo, D Marzorati, D Colombo, A Ferretti, A Bagliani, and F Bottazzi. Geomechanical response to seasonal gas storage in depleted reservoirs: A case study in the po river basin, italy. *Journal of Geophysical Research: Earth Surface*, 116(F2):n/a–n/a, 2011.

[17] M D McKay, W J Conover, and R J Beckman. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. 21(2), 1979.

[18] M E Johnson, L M Moore, and D Ylvisaker. Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, 26(2):131–148, 1990.

[19] A Lovison and E Rigoni. Extracting optimal datasets for metamodelling and perspectives for incremental samplings. *Mathematics and Computers in Simulation*, 81(3):681–692, November 2010.