A Fresh Look at Variography: Measuring Dependence and Possible Sensitivities Across Geophysical Systems From Any Given Data

Razi Sheikholeslami1,2,3 and Saman Razavi2,3,4

1Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford, UK, 2School of Environment and Sustainability, University of Saskatchewan, Saskatoon, Saskatchewan, Canada, 3Global Institute for Water Security, University of Saskatchewan, Saskatoon, Saskatchewan, Canada, 4Department of Civil, Geological, and Environmental Engineering, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

Abstract Sensitivity analysis in Earth and environmental systems modeling typically demands an onerous computational cost. This issue coexists with the reliance of these algorithms on ad hoc designs of experiments, which hampers making the most out of the existing data sets. We tackle this problem by introducing a method for sensitivity analysis, based on the theory of variogram analysis of response surfaces (VARS), that works on any sample of input-output data or pre-computed model evaluations. Called data-driven VARS (D-VARS), this method characterizes the relationship strength between inputs and outputs by investigating their covariograms. We also propose a method to assess “robustness” of the results of using a pre-existing set of input-output data or pre-computed model evaluations. Using two hydrologic modeling case studies, we show that D-VARS is highly efficient and statistically robust, even when the sample size is small. Therefore, D-VARS can provide unique opportunities to investigate geophysical systems whose models are computationally expensive or available data is scarce.

Plain Language Summary Sensitivity analysis (SA) is about assessing how the properties of a system are influenced by different factors. It can also help us better understand the behavior of a mathematical model and the underlying real-world system that it mimics. Almost always, classic SA estimates the sensitivities by sampling the entire problem space in a specific manner. They are incapable of using a pre-existing set of input-output data or pre-computed model evaluations. Hence, classic SA becomes useless in cases where a sample of input-output data, obtained from physical experiments or computationally expensive simulations, is already available. We propose a purely data-driven method that can effectively be used in such situations. Based on the principles of variography, our method measures dependence and possible sensitivities across a system from any given data. Here, we illustrate our method in the context of hydrologic modeling, but it can potentially be applied to study other geophysical systems models.

1. Introduction

Understanding how various uncertain factors influence Earth and environmental systems (EES) models’ behavior has greatly raised the need for continued development of the efficient methodologies for global sensitivity analysis (WSA). GSA methods identify dominant factors that exert a considerable impact on the model responses and accordingly provide invaluable information for model simplification, risk assessment, and uncertainty analysis (Bhalachandran et al., 2019; Castelletti et al., 2012; Janetti et al., 2019; Li et al., 2019; Markstrom et al., 2016; Puy et al., 2020; Safta et al., 2015).

Almost all GSA methods are “sampling-based,” starting by sampling from a d-dimensional factor/input space using design of experiments approaches. Next, the corresponding response variable of interest at all sample points are determined from an output space through an input-output relationship, that is, a model. A statistical estimator can then be employed to compute the sensitivity indices, which are essentially a representation of relationship strength between the output and individual inputs. However, two major issues preclude efficient application of the sampling-based GSA:

1. They are bound to their own sampling strategies that follow specific spatial arrangement (the “ad hoc designs” as termed by Ratto et al., 2007).
(2) They are not applicable when the underlying input-output functional relationship is unavailable (e.g., we may only have a sample of input-output data obtained from field observations or previous modeling experiments, and nothing more).

The former issue is widespread among GSA methods (e.g., GSA methods of Saltelli et al., 2010, and Razavi & Gupta, 2016a, 2016b), and therefore, a sample taken by one method cannot be utilized by another method. Furthermore, this complicates comparison of the different methods. And due to the latter, GSA has remained limited to cases where a computational model is available, and its run time is tractable to generate a large enough input-output sample. Hence, the mainstream methods are not useful when the distributions, correlations, and interactions between different factors need to be characterized based on the existing data set, without model (re)evaluations. These challenges necessitate a “given-data approach” to GSA (alternatively termed as “data-driven approach” or “post-processing GSA,” see Borgonovo et al., 2016; Li & Mahadevan, 2016; Plischke, 2010) that can extract sensitivity-related information from preexistent data sets. In a burgeoning era of big data, this becomes more crucial than ever because data collection can be much easier and faster than building models.

A limited number of GSA methods can potentially be used under the given-data approach. These are mainly emulator-based techniques, limited to estimating variance-based sensitivity indices using either Monte Carlo-based method (see, e.g., Iooss et al., 2006; Marseguerra et al., 2003; Storlie et al., 2009) or analytical approaches (see, e.g., Jia & Tatianidis, 2016; Jourdan, 2012; Sargsyan & Safta, 2019; Sudret, 2008). Additionally, moment-independent methods can also be used in the given-data setting (Borgonovo et al., 2012, 2017; Plischke et al., 2013; Yun et al., 2018). Each of these methods characterizes only a specific feature of the response surface and ignores the rest (Razavi & Gupta, 2015). Recognizing this fact, some researchers have recently advocated the use of multiple GSA measures together, which may be based in different theories, at the cost of augmenting computational burden (Borgonovo et al., 2017; Pianosi et al., 2016; Wang & Solomatine, 2019). We should note that, in theory, the given-data paradigm provides a unique platform to estimate various GSA measures simultaneously from the same data set, without much additional computational effort.

Here, we contribute to the given-data paradigm by developing a highly efficient and robust data-driven GSA, based on the theory of variogram analysis of response surfaces (VARS) (Razavi & Gupta, 2016a, 2016b). Despite the increasing popularity of VARS (e.g., Akomeah et al., 2019; Becker, 2020; Korgaonkar et al., 2020; Krogh et al., 2017; Leroux & Pomeroy, 2019; Lilhare et al., 2020; Schürz et al., 2019; Sheikholeslami et al., 2017; Yassin et al., 2017), its current version cannot be applied in the given-data setting. Conversely, our new version of VARS works on any given data, by approximating the anisotropic variogram structure of the underlying (but unknown) response surface, when only a (small) sample of the input-output space is available. We also address a crucial, but often neglected, component of any GSA practice, which is assessing the “robustness” of its results. Note that GSA results are typically prone to statistical uncertainty due to sampling variability and to imperfect nature of the numerical methods. However, a comprehensive robustness assessment can be computationally costly or infeasible in the given-data context. Therefore, we develop a new robustness index that works well within the given data setting.

2. Method

2.1. Background

VARS is a new GSA framework that builds on anisotropic variograms to quantify the influence of input factors on the variability of response variables. Directional variograms are a rich source of information to attribute the structure and spatial variability of a response variable to the distributional properties of different factors. VARS recognizes that the variability of any continuous response surface can be better expressed by the variance of change in the response variable, when a factor or group of factors are perturbed with varying perturbation sizes across the factor space. In other words, for any pairs of sample points in the factor space, for example, \( X^a \) and \( X^u \), the variance of the difference between the corresponding response variables, \( y(X^a) \) and \( y(X^u) \), depends on their distance \( ||X^a - X^u|| = h \) in the \( d \)-dimensional input space \( \mathbb{R}^d \), that is
\[
\text{var}[y(X^u) - y(X^w)] = E[(y(X^u) - y(X^w))^2] = 2\gamma(h) \quad u, w = 1, 2, ..., m
\] (1)

where \( h = [h_1, h_2, ..., h_d] \) denotes the vector that separates two sample points with respect to distance and direction and \( \gamma(h) \) is one half of the expected squared difference between \( y(X^u) \) and \( y(X^w) \).

If the stationarity assumption holds (Matheron, 1971), the function that relates \( \gamma \) to \( h \), known as (semi)\( \gamma \) 

\[
\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (y(X^i) - y(X^i + h))^2
\] (2)

where \( N(h) \) is the number of all pairs of the sample points in the input space separated by the distance vector \( h \). The vector \( h \) is also referred to as “perturbation scale” in VARS terminology. An example of a response surface and the estimated variogram surface is shown in Figure S1 in the supporting information.

Using Equation 2, directional variograms along the \( j \)th input factor (\( j = 1, 2, ..., d \)) can be estimated by calculating \( \gamma(h) \) for any given set of \( h_j \), for example, \([0, \Delta h, 2\Delta h, ...] \) with perturbation resolution of \( \Delta h \) (Razavi & Gupta, 2016b). The directional variograms, therefore, show how the variability of the response variable is changing with respect to the direction and perturbation scale. Accordingly, VARS-based sensitivity analysis links the rate of variability to both direction and perturbation scale (for detail see Razavi & Gupta, 2016a).

Finally, to measure factor sensitivities, given perturbation scales ranging from 0 to \( H_j \), VARS defines a set of sensitivity indices for the \( j \)th factor, called integrated variograms across a range of scales (IVARS), as follows

\[
\Gamma(H_j) = \int_0^{H_j} \gamma(h_j) \, dh_j
\] (3)

2.2. New Extension of VARS: A Given-Data Estimator

We extend the theory of VARS and propose a data-driven estimator, called D-VARS. The stationarity assumption (Matheron, 1971) implies that the variogram can be related to spatial covariance of the response variable according to the following equation (see Text S1 in the supporting information)

\[
\gamma(h) = \text{var}[y] - \text{cov}[y(X^u), y(X^w)] = \sigma^2 - C(h)
\] (4)

where \( \sigma^2 \) and \( C(h) \) are the variance and spatial covariance function of the \( y(X) \), respectively.

Hence, for any given covariance function, the variogram of the response variable is uniquely constructible from the covariance function, which must be a symmetric, positive definite function of \( h \) (Rasmussen & Williams, 2006). To characterize the covariance functions in D-VARS, we assume a zero mean Gaussian process (GP) throughout this study. In the case of GP, the covariance function can be written as (Jones, 2001; Sacks et al., 1989)

\[
C(h) = \sigma^2 R(h)
\] (5)

where \( R(\cdot) \) is the correlation function that provides spatial correlation properties.

We focus on correlation functions that can be defined as a product of one-dimensional kernel functions, that is, \( r_j \):

\[
R(X^u, X^w) = R(h) = \prod_{j=1}^{d} r_j(h_j), \quad j = 1, 2, ..., d
\] (6)

Following Equations 4–6, the variogram can be obtained by

\[
\gamma(h) = \sigma^2 (1 - R(h)) = \sigma^2 \left( 1 - \prod_{j=1}^{d} r_j(h_j) \right)
\] (7)

By substituting Equation 7 in Equation 3, D-VARS estimates the IVARS sensitivity indices for any perturbation scale from zero to \( H_j \) (for the \( j \)th factor), as follows:
Theoretically, the one-dimensional correlation functions, \( r_j \), can be learned purely from input-output data. With covariance functions distilled from data, D-VARS directly estimates several sensitivity indices using Equation 8. We discussed numerical implementation of D-VARS in Text S2 of the supporting information. Moreover, inspired by Sheikholeslami et al. (2019), we defined a robustness index (see Text S2) to evaluate robustness of D-VARS.

3. Numerical Experiments

3.1. Computer Experiments Versus Physical Experiments

We argue that D-VARS is applicable to both computer experiments and physical experiments. For the former, D-VARS contributes to more efficient and sampling-free GSA, while for the latter, it provides new opportunities to assess the strength of relationships, either causal or correlational, between different variables measured in an experiment. There are two major differences between these two types of experiments: computer experiments are usually deterministic, while data collected from physical experiments are prone to noise or errors, often with unknown properties. In computer experiments, one may have full control on the experimental design for collecting samples and their distributional properties, while it is typically not the case in physical experiments. In this paper, we chose to test D-VARS on computer experiments. While the outcome of our tests here can, in part, be generalized to physical experiments, a rigorous study is required to test D-VARS to cases wherein data sets are polluted with noise and variables follow a variety of distributional forms. This will be the topic of our follow-up paper.

3.2. Hydrologic Models Used

We designed our case studies with two well-established hydrologic models of increasing complexity to illustrate the utility of D-VARS. The HYMOD model with five parameters was configured for the 1,950-km² Leaf River catchment located in Mississippi, USA. The HBV-SASK model with 12 parameters was configured for the 1,435-km² Oldman River basin located in Alberta, Canada. The model parameters were chosen as inputs to D-VARS. As the output, for the HYMOD case study, we chose a goodness-of-fit metric to observations, Nash-Sutcliffe efficiency (NS) (Nash & Sutcliffe, 1970), as a common setting that informs calibration (Gupta & Razavi, 2018). For the HBV-SASK case study, however, we chose model’s estimate of flood peak with the 10-year return period, as a more modern application of GSA focused on learning about the system’s behavior under uncertainty and nonstationarity (Razavi et al., 2019, 2020). Full details are available in the supporting information (see Text S3).

3.3. D-VARS Runs

We first ran the original VARS for both case studies using the sampling-based STAR-VARS algorithm (Razavi & Gupta, 2016b). A large sample size was chosen, resulting in ~70,000 model evaluations for each case study, to ensure convergence to robust and accurate results. These results were deemed to be the “true” sensitivities and used as the comparison benchmark for the performance assessment of D-VARS.

Second, we generated synthetic data sets by randomly sampling from the input space, with progressively increasing sample size (i.e., 20, 50, 100, 200, and 400 sample points) via progressive Latin hypercube sampling (Sheikholeslami & Razavi, 2017). We then ran all the sample sets through the models to obtain the respective outputs. Each input-output sample set was considered to be a set of “given data.” We replicated this data generation process 100 times with different initial random seeds to assess both average and variability of the D-VARS behavior. Furthermore, IVARS-50, called “total-variogram effect” (Razavi et al., 2019), was applied to assess the factor sensitivities, which means that Equation 8 was computed for \( H_j = 50\% \) of the parameter range.

4. Results and Discussion

4.1. Factor Sensitivities and Actual Robustness of D-VARS

Figures 1 and 2 present the histograms of the parameter rankings for different input-output sample sizes. Rank1 stands for the most influential parameter on the output, i.e., the dependency of the output to the
Rank1 parameter is the highest. The true ranking of the HYMOD and HBV-SASK parameters are as follows: \{R_q, C_{max}, \alpha, b_{exp}, and R_s\} (see Table S1) and \{PM, C0, FRAC, TT, FC, a, UBAS, K1, EFT, LP, K2, and beta\} (see Table S2) from the most influential to the least influential one.

As shown in Figures 1 and 2, rankings of the most influential parameters for HYMOD, \{R_q, C_{max}, and \alpha\}, and HBV-SASK, \{PM, C0, FRAC, TT, and FC\}, have been well established even when the size of given data was 50. This confirms that D-VARS is extremely efficient in identifying the most influential parameters, when the sample size is very small. Furthermore, Figure 2 shows that the ranking of parameters with moderate importance in HBV-SASK, \{\alpha, UBAS, and K1\}, stabilized when the sample size was larger than 200. A close examination of the results reveals that in more than 50 replicates, the ranking of these parameters converged to the true ranking, when the sample size is only 200. It is noteworthy that parameters with the least influence on HBV-SASK, \{EFT, LP, K2, and beta\}, have the slowest convergence rate in terms of ranking.

Now let us directly look at actual sensitivity indices, IVARS-50, derived by D-VARS. Figures 3 and 4 show IVARS-50 values (scaled between 0 and 1) for the HYMOD and HBV-SASK parameters, obtained from the 100 replicates of each experiment. For an extremely small sample size (i.e., 20), D-VARS showed highly variable performance. However, by increasing the sample size, all the replicates quickly converged to a single set of IVARS-50 (i.e., true values), particularly for the most influential parameters. This confirms the
robustness of D-VARS against sampling variability. For the least influential parameters, the IVARS-50 values may not be distinguishable across all the replicates, even for the larger sample sizes. This is mainly because these parameters are almost non-influential on the output of interest and their associated IVARS-50 values are close to 0.

4.2. Physical Justifiability of Sensitivity Assessments

A critical question that one might ask after running GSA is whether the results are physically meaningful. Based on our results, we can (rather subjectively) categorize the parameters of HYMOD into two groups: influential: \([R_q, C_{max}, \text{ and } \alpha]\), and non-influential: \([b_{exp} \text{ and } R_s]\), and those of HBV-SASK into three groups: (i) strongly influential: \([PM, C_0, \text{ FRAC, TT, and FC}]\), (ii) moderately influential: \([a, UBAS, \text{ and } K_1]\), and (ii) non-influential: \([EFT, LP, K_2, \text{ and } \beta]\). Recall that these assessments are based on choosing NS values and 10-year flood estimates as the outputs for the HYMOD and HBV-SASK case studies, respectively. We know from hydrology domain knowledge that both outputs should, in principle, be dominantly controlled by high flows in the hydrograph.

In case of HYMOD, D-VARS ranked \(R_q\), a parameter controlling the quick flow generation, as the most influential parameter. This assessment is physically justifiable, as the Leaf River basin is a rainfall-dominated basin with a history of torrential storms. Most influential parameters of HBV-SASK,
however, are those mainly responsible for the snowmelt (\(C_0\) and \(TT\)) and soil (\(FRAC\) and \(FC\)) processes. The high influence of \(C_0\) and \(TT\) can be justified because in the Oldman River basin major floods are governed by snow accumulation and melt in early spring. From the structure of both models, it is evident that \(\alpha\) (in HYMOD) and \(FRAC\) (in HBV‐SASK) determine the fraction of soil release entering fast reservoir and accordingly play a significant role on the high flow values. Moreover, \(C_{\text{max}}\) (in HYMOD) and \(FC\) (in HBV‐SASK) account for partitioning of the precipitation into runoff and soil moisture, and thus can significantly impact the simulated high flows. Additionally, our method recognized \(\alpha\) and \(UBAS\) among the influential parameters for peak flow generation in HBV‐SASK since they control timing and attenuation of the release from the fast reservoir. Finally, D‐VARS identified \(R_s\) (in HYMOD) and \(K_2\) (in HBV‐SASK) as the non‐influential parameters. These parameters represent the release pace of slow reservoir in the structure of these models, and as such are only responsible for base flows (minimally contributing to peak flows).

4.3. The Proposed Robustness Index

The assessment of actual robustness, as presented in section 4.1, is typically infeasible in practice because only a single data set is often available or the model under investigation is computationally expensive to generate multiple independent sets of input‐output data. This study, therefore, presented a novel robustness index that can estimate the robustness of D‐VARS for any given data (see Text S2). The performance of this new index for the synthetic samples taken from the HYMOD and HBV‐SASK models across 100 replicates are depicted in Figure 5.

As shown in Figure 5, by increasing the sample size, the variability of the robustness index obtained over all replicates of the algorithm became lower, and consequently the robustness of D‐VARS became higher. Also, for both models, the robustness indices quickly converged toward 100% (i.e., perfect robustness). When the sample size was larger than 100, almost all the robustness indices were higher than 50%. For clarity, see the

Figure 3. Sensitivity indices for the HYMOD parameters obtained by D‐VARS. Subplots (a–e) show the IVARS-50 values from given data with different sample sizes for 100 replicates of the experiment.
medians of the estimated robustness indices (the horizontal black lines in Figure 5). Notice that the medians of the robustness indices were already quite high when the sample size was 100 (55% for HYMOD and 79% for HBV-SASK) and increased rapidly thereafter to be 90% for HYMOD and 93.7% for HBV-SASK at 400 samples.

Figure 4. Sensitivity indices for the HBV-SASK parameters obtained by D-VARS. Subplots (a–e) show the IVARS-50 values from given data with different sample sizes for 100 replicates of the experiment.

Figure 5. Robustness assessment results for the (a) HYMOD and (b) HBV-SASK models. Boxplots represent the distribution of robustness index across the 100 replicates. Robustness index varies between 0% to 100%, with the latter corresponding to perfectly robust results.
5. Conclusions

GSA has often been tied to ad hoc experimental designs and defined in the context of mathematical models. This study tried to take GSA to a next level by proposing a new method to conduct GSA for any given data, independent of experimental designs and mathematical models. Our proposed method, called D-VARS, was based on variography via GP modeling to characterize the spatial correlation properties of the underlying response surface for estimating sensitivity/dependence indices. D-VARS is not only a computationally cheap method that works with any given data but also makes, in principle, less confining assumptions than most of the existing sampling-based methods. For example, many GSA methods assume that the input data themselves are uncorrelated (see Do & Razavi, 2020), while D-VARS handles correlated inputs as well (not shown in this paper). We examined the performance of the method using two benchmark hydrologic case studies. Overall, our results demonstrated that D-VARS accurately estimates the true sensitivity measures requiring very small sample sizes.

The effectiveness and high-efficiency of D-VARS make it uniquely positioned to advance GSA paradigm on two fronts:

(1) D-VARS can open up a new area of research, where GSA is applied to any data set, even when the underlying relationship and mechanisms are not known.

(2) D-VARS can enable GSA for computationally expensive models, wherein conventional GSA is handicapped, as D-VARS requires minimal computational effort to produce robust results.

Further, we argued that any GSA practice must be accompanied by the assessment of the robustness, which is typically neglected in the literature. To this end, we developed a new robustness index and showed that it can consistently provide an accurate evaluation of robustness. Our proposed index can be easily used in practice since it works when only a single (albeit small) data set is available. Future work may include testing our proposed method on real-world data obtained from field observations or remote sensing to better support model development and understanding.

Data Availability Statement

Data sets for this research is available through Razavi et al. (2019).

References

Akoura, E., Lindenschmidt, K. E., & Chapra, S. C. (2019). Comparison of aquatic ecosystem functioning between eutrophic and hypereutrophic cold-region river-lake systems. Ecological Modelling, 393, 25–36. https://doi.org/10.1016/j.ecolmodel.2018.12.004

Borgonovo, E. (2020). Metafunctions for benchmarking in sensitivity analysis. Reliability Engineering & System Safety, 204, 107189. https://doi.org/10.1016/j.ress.2020.107189

Bhalachandran, S., Haddad, Z. S., Hristova-Veleva, S. M., & Marks, F. D., Jr. (2019). The relative importance of factors influencing tropical cyclone rapid intensity changes. Geophysical Research Letters, 46, 2282–2292. https://doi.org/10.1029/2018GL079997

Borgonovo, E., Castaings, W., & Tarantola, S. (2012). Model emulation and moment-independent sensitivity analysis: An application to environmental modelling. Environmental Modelling & Software, 34, 105–115. https://doi.org/10.1016/j.envsoft.2011.06.006

Borgonovo, E., Haen, G. B., & Plischke, E. (2016). A common rationale for global sensitivity measures and their estimation. Risk Analysis, 36(10), 1871–1895. https://doi.org/10.1111/risa.12555

Borgonovo, E., Lu, X., Plischke, E., Rakovec, O., & Hill, M. C. (2017). Making the most out of a hydrological model data set: Sensitivity analyses to open the model black-box. Water Resources Research, 53, 7933–7950. https://doi.org/10.1002/2017WR020767

Castelletti, A., Galelli, S., Ratto, M., Soncini-Sessa, R., & Young, P. C. (2012). A general framework for dynamic emulation modelling in environmental problems. Environmental Modelling & Software, 34, 5–18. https://doi.org/10.1016/j.envsoft.2012.01.002

Do, N. C., & Razavi, S. (2020). Correlation effects? A major but often neglected component in sensitivity and uncertainty analysis. Water Resources Research, 56, e2020WR025436. https://doi.org/10.1029/2020WR025436

Gupta, H. V., & Razavi, S. (2018). Revisiting the basis of sensitivity analysis for dynamical Earth system models. Water Resources Research, 54, 8692–8717. https://doi.org/10.1029/2018WR026668

Iooos, B., Van Dorpe, F., & Devictor, N. (2006). Response surfaces and sensitivity analyses for an environmental model of dose calculations. Reliability Engineering & System Safety, 91(10–11), 1241–1251. https://doi.org/10.1016/j.ress.2005.11.021

Janetti, E. B., Guadagnini, L., Rita, M., & Guadagnini, A. (2019). Global sensitivity analyses of multiple conceptual models with uncertain parameters driving groundwater flow in a regional-scale sedimentary aquifer. Journal of Hydrology, 574, 544–556. https://doi.org/10.1016/j.jhydrol.2019.04.035

Jia, G., & Taitlanidis, A. A. (2016). Efficient evaluation of Sobol’ indices utilizing samples from an auxiliary probability density function. Journal of Engineering Mechanics, 142(5), 04016012. https://doi.org/10.1061/(ASCE)EM.1943-7889.0001061

Jones, D. R. (2001). A taxonomy of global optimization methods based on response surfaces. Journal of Global Optimization, 21(4), 345–383. https://doi.org/10.1023/A:1012771025575

Jourdan, A. (2012). Global sensitivity analysis using complex linear models. Statistics and Computing, 22(3), 823–831. https://doi.org/10.1007/s11222-011-9239-y
Wang, A., & Solomatine, D. P. (2019). Practical experience of sensitivity analysis: Comparing six methods, on three hydrological models, with three performance criteria. Water, 11(5), 1062. https://doi.org/10.3390/w11051062

Yassin, F., Razavi, S., Wheeler, H., Sajiri-Azari, G., Davison, B., & Pietroniro, A. (2017). Enhanced identification of a hydrologic model using streamflow and satellite water storage data: A multicriteria sensitivity analysis and optimization approach. Hydrological Processes, 31(19), 3320–3333. https://doi.org/10.1002/hyp.11267

Yun, W., Lu, Z., Jiang, X., & Zhang, L. (2018). Borgonovo moment independent global sensitivity analysis by Gaussian radial basis function meta-model. Applied Mathematical Modelling, 54, 378–392. https://doi.org/10.1016/j.apm.2017.09.048

References From the Supporting Information

Bergström, S. (1995). The HBV model. In V. P. Sing (Ed.), Computer models of watershed hydrology (pp. 443–476). Highlands Ranch, CO: Water Resources Publ.

Boggs, P. T., & Tolle, J. W. (2000). Sequential quadratic programming for large-scale nonlinear optimization. Journal of Computational and Applied Mathematics, 124(1–2), 123–137. https://doi.org/10.1016/S0377-0427(00)00429-5

Boyle, D. P., Gupta, H. V., & Sorooshian, S. (2000). Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. Water Resources Research, 36(12), 3663–3674. https://doi.org/10.1029/2000WR900207

Butler, A., Haynes, R. D., Humphries, T. D., & Ranjan, P. (2014). Efficient optimization of the likelihood function in Gaussian process modelling. Computational Statistics & Data Analysis, 73, 40–52. https://doi.org/10.1016/j.csda.2013.11.017

Chilès, J.-P., & Delfiner, P. (1999). Geostatistics: Modeling spatial uncertainty. New-York: Wiley. https://doi.org/10.1002/9780470316993

Efron, B. (1982). The jackknife, the bootstrap and other resampling plans. Philadelphia: SIAM.

Liu, H., Ong, Y. S., Shen, X., & Cai, J. (2020). When Gaussian process meets big data: A review of scalable GPs. IEEE Transactions on Neural Networks and Learning Systems, 1–19. https://doi.org/10.1109/TNNLS.2019.2957109

Lophaven, S. N., Nielsen, H. B., & Søndergaard, J. (2002). Aspects of the Matlab toolbox DACE. In Technical report IMM-REP-2002-13. Denmark: Technical University of.

MacDonald, B., Ranjan, P., & Chipman, H. (2015). GPlab: An R package for fitting a Gaussian process model to deterministic simulator outputs. Journal of Statistical Software, 64(12), 1–23. https://doi.org/10.18637/jss.v064.i12

Powell, M. J. D. (1978). A fast algorithms for nonlinearly constrained optimization calculations. In G. A. Watson (Ed.), Numerical analysis (pp. 144–157). Berlin: Springer-Verlag. https://doi.org/10.1007/BFb0067703

Regis, R. G. (2016). Trust regions in Kriging-based optimization with expected improvement. Engineering Optimization, 48(6), 1037–1059. https://doi.org/10.1080/03092170.2015.1082350

Regis, R. G., & Shoemaker, C. A. (2013). A quasi-multistart framework for global optimization of expensive functions using response surface models. Journal of Global Optimization, 56(4), 1719–1753. https://doi.org/10.1007/s10898-012-9940-1

Schittkowski, K. (1983). On the convergence of a sequential quadratic programming method with an augmented Lagrangian line search function. Mathematische Operationsforschung and Statistik. Series Optimization, 14(2), 197–216.

Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., & Xu, C. (2015). Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. Journal of Hydrology, 522, 739–757. https://doi.org/10.1016/j.jhydrol.2015.02.013

Vrugt, J. A., Gupta, H. V., Bouten, E., & Sorooshian, S. (2003). A shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resources Research, 39(8), 1201. https://doi.org/10.1029/2002WR001642

Wagener, T., Boyle, D. P., Lees, M. J., Wheeler, H. S., Gupta, H. V., & Sorooshian, S. (2001). A framework for development and application of hydrological models. Hydrology & Earth System Sciences, 5(1), 13–26. https://doi.org/10.5194/hess-5-13-2001

Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J., & Morris, M. D. (1992). Screening, predicting, and computer experiments. Technometrics, 34(1), 15–25. https://doi.org/10.2307/1269548