1. Introduction

Regional sensitivity analysis, RSA, is now approaching its 40th birthday (Spear & Hornberger, 1980). Not surprisingly, over the years, it has seen various elaborations and sophistications of the original procedure, notably the Generalized Likelihood Uncertainty Estimation methodology (Beven & Binley, 1992). RSA has also been referred to using other descriptions, for example, “sensitivity tests involving segmented input distributions” (Hamby, 1994) and Monte Carlo filtering (Saltelli et al., 2008). From a statistical perspective, RSA was recognized early in its history as an essentially Bayesian approach to the incorporation of prior information on parameter values and system behavior into model-based analyses (Sharefkin, 1983). Later, Poole and Raftery (2000) provided a more formal statistical structure that they termed Bayesian Melding in which the two types of prior information are melded to inform a posterior distribution of the model parameters. RSA generates random samples from this complex posterior distribution. Despite the various names by which RSA may be known or the complexities of its underlying statistical structure, the central concept and the original procedure remain appealing because of their inherent simplicity. In that spirit, this paper reports an exploration of augmenting the original approach, first to address including interactions in the ranking of parametric importance using machine learning methods (MLMs) available as general purpose open-source software. Further, we report the application of a related density estimation technique that provides some insight into the posterior distribution of parameters mentioned above as well as the importance of the scale of sensitivity in RSA and to sensitivity analyses generally. Density estimation illustrates the concept of equifinality, an implicit result of RSA that has received inadequate attention in environmental modeling generally with a notable exception being in the field of hydrology (Beven & Freer, 2001). In that context, a recent paper contains a useful review of a number of other applications of MLMs in hydrology (Teweldebrhan et al., 2019).

2. Regional Sensitivity Analysis

Although hydrology has been the field in which the majority of the RSA applications have been reported, prior to introducing an example of augmenting the standard use of RSA using MLMs, we begin with a
summary of its essential features adapted from an early publication (Hornberger & Spear, 1980). In common with the model addressed in this report, assume generally that the model to be analyzed is described by a set of ordinary nonlinear differential equations (different mathematical structures can be dealt with in an analogous way):

$$\frac{dx(t)}{dt} = f[x(t), \xi, z(t)],$$

where $x(t)$ is the state vector and $z(t)$ a set of time variable functions which include input or forcing functions. The vector $\xi$ is comprised of constant parameters described more fully below. Thus, for $z(t)$ and $x(0)$ specified, $x(t)$ is the solution of the system and is a deterministic or a stochastic function of time as determined by the nature of $z(t)$. Each element of the vector $\xi$ is defined as a random variable, the distribution of which is a measure of our uncertainty in the real but unknown value of the parameter. These parameter distributions are formed from data available from the literature and from experience with similar processes applied in other settings. These a priori estimates of the elements of $\xi$ result in the definition of an ensemble of models. Some members of this ensemble will, we hope, mimic the real system with respect to its behavior of interest. In particular, we assume there are observable characteristics of the output of the model, $y(t) = g[x(t)]$, that can be used to compare with the field data. Unlike MLMs, which require large data sets, mathematical models rely heavily on assumptions and mechanisms and can be used for small data sets as well.

The behavior of interest in the original application of RSA concerned the nature and amount of data available for calibration of a model of an eutrophic inlet in Western Australia (Hornberger & Spear, 1980). The principal question in that analysis concerned identifying processes within the system that would benefit from further field study. Because the field data defining the behavior of the system were sparse and subject to uncertainty, only its pattern was believed sufficiently robust to define the characteristic behavior of concern. A binary classification was applied to the model output, that is, the pattern of the output either adequately matched that of the field data or it did not, a pass or a fail. This classification allowed for diverse forms of pass/fail criteria that depended, for example, on the timing of events or conditional criteria. It should be noted that the pass/fail binary criterion used in the early applications, and in much of the subsequent work, can be easily extended from two, pass/fail, to multiple classes as proposed, for example, by Fenwick et al. (2014) and implemented in the DGSA Matlab toolbox (Park et al., 2016). Three classes might, for example, be based on meeting all, a subset, or none of the elements of the pass/fail criteria. In this context, early in any application of RSA, it is wise to assess the stringency of the individual elements of the pass criteria because there can be strong negative correlations among them that may reflect unanticipated structural constraints of the model or unintended emphasis on a particular characteristic of the system behavior.

The simulation procedure is simply to select a random parameter vector from prior distributions on $\xi$, run the model, and then test the output to classify it as leading to a pass or a fail. The set of passing samples are random samples from the posterior distribution of the melded priors in the context of Bayesian melding. The novelty of the procedure was then to explore the degree to which the statistical distribution of passing parameter vectors differed from that of failing vectors. The magnitude of this difference was regarded as a measure of parametric sensitivity. While it has been argued that the term sensitivity currently has a much broader and rather different meaning from that used here (Gupta & Razavi, 2018), we will adhere to its traditional meaning in the limited sense that our interest is in model calibration or in identifying processes within the model that may explain the observed behavior of the system and/or may point to interventions for modifying its behavior.

The original, and most subsequent, applications of RSA have used the univariate marginal distributions of the elements of $\xi$ to assess the difference between the prior parameter distribution and that of the melded posterior, often applying the Kolmogorov-Smirnov two-sample test to the cumulative distributions of each set of marginals. However, it has been clear from the outset that parameter importance ranking can also be affected by parameter interactions that may not be reflected by the univariate marginals. Here, we explore MLMs as a means of addressing this deficiency. The original RSA process as well as the extension using MLMs will be illustrated using the mechanistic model of dengue transmission described below.
3. The Dengue Model and Standard RSA

RSA and its augmented versions are demonstrated in the context of a previously studied mechanistic model of dengue transmission. The model was developed to examine the drivers of a 2014 dengue epidemic and was comprised of 11 differential equations (state variables) and 19 parameters (Cheng et al., 2017). The model simulates the life cycle of *Aedes albopictus* mosquitoes—the sole disease vector of dengue in our study area—under different climate conditions and water availability, as well as the transmission of dengue virus between human and mosquitoes depending on the contact between them through mosquito biting. The schematic of the model and the equations used are presented in Figure S1 and Text S1 in the supporting information.

The prior distribution of model parameters was generated by reviewing prior literature or relying on best estimates from local sources (Table 1). Because the standard RSA procedure requires classification of each sampled parameter vector as a pass or fail, we define a pass to be any parameter vector that produces a daily count of incident dengue cases which matches observed data in both timing and magnitude (Figure S2). The temporal trends of the amount of larvae and adults generated by these passing parameter vectors also match local mosquito surveillance data, which further supports the validity of our model (Figure S3). More confidence was placed on the timing and pattern of the calibration data than in the precise numbers reported on any day as is typical in applications of the RSA procedure. Given the model structure, the parameter distributions, and the calibration data, the next step was the repetitive simulation of the model, each simulation using a randomly selected parameter vector from the prior distributions. Parameter values sampled from the prior distributions and their corresponding pass/fail status were stored.

In common with other RSA applications, the next step in the dengue model analysis was to “trim” the univariate marginals to raise the pass rate and lessen the computing cost of gaining sufficient passes for more detailed analysis. Although it has been our experience with the RSA procedure that passing parameter vectors generally extend over nearly the entire prior distributions of the individual parameters, it is common for few passes to be observed near the bounds of some parameters. This can be seen in the cumulative density curves of passes and fails for each parameter of the dengue model that are plotted and compared in Figures S4 and S5. Ranges with flat cumulative density curves (usually less than 5% of all passes) were trimmed for the next round of simulation. For example, the range of $\beta_{2013}$ was reduced from 521–571 to 555–570 because only two out of 83 passes have a $\beta_{2013}$ in the trimmed range (Figures S4). Clearly, the trimming step using only the univariate marginal distributions is a pragmatic approach which ignores all parameter correlations of any order. However, when the information on the pass space is based principally on the univariate or bivariate marginals, it has been a useful step. Nevertheless, virtually any reduction in the prior parameter space will exclude some number of passing simulations that met the specified definition of an acceptable fit to the observed behavior of the model. The implications of such exclusions are application specific but need to be considered in interpretations of the final results.

Following relatively modest trimming of nine of the 19 marginal parameter distributions (Table S1), informed by the first (Cycle 1) set of 800,000 runs and 83 passes, the second set (Cycle 2) of 200,000 runs resulted in 948 passes in the new search space with 1.47% of its original volume in Cycle 1. To avoid class imbalance, which can lead to anomalies in some MLM algorithms, 948 fails were randomly sampled from the 199,052 fails for training the subsequent MLMs (Japkowicz, 2000). These passing and failing vectors were used first for ranking relative parameter importance and then for mapping the parameter space, which leads then to issues of scale and screening.

Given the binary classifier, the original RSA application relied principally on comparisons of the univariate marginal distributions of each parameter classified as passing or failing the classification criteria rather than dealing with the multivariate parameter distribution itself. Originally, the Kolmogorov-Smirnov statistic, $d_m$, was used to assess the difference between the univariate sample distribution functions of passing and failing parameter vectors with its value used for both ranking and screening (see section 4 for the definition of ranking and screening). From the outset, it was recognized that there were almost certainly moderate degrees of correlation, at least locally, between some parameters in the high-dimensional posterior space of passing vectors. Hence, exploring the univariate distributions was understood to be only the first step in assessing parameter sensitivity in the high-dimensional spaces typical of this class of models, typically...
20 or more in many published applications. At the time, however, to assess interactions beyond two-by-two correlations was difficult (Beck, 1987).

Figure S5 shows the univariate marginal distributions from Cycle 2. At this scale, there are five parameters that result in very small differences in the univariate marginal between passes and fails, $\theta$, $\lambda$, $\tau_{\text{exh}}$, $\tau_{\text{ih}}$, and $\delta$. Further, none of these five parameters shows two-by-two correlations above 0.23 with any other parameter (Figure S6). The original RSA analysis and many subsequent applications would stop at this point and conclude that these five parameters were insensitive, and further analysis would focus on the 14 remaining parameters. However, interactions invisible from the marginal cannot be discounted nor can issues relating to the scale of the Cycle 2 analysis as addressed below.

An obvious means of exploring parameter interactions that might not be visible from the univariate marginals is to explore pass/fail differences in lower dimensional subspaces that are more amenable to visualization. Because most applications of RSA are to the analysis of mechanistic mathematical models, the explicit modeling of physical processes often suggests important and interesting interactions, the importance of which may wax or wane given prior uncertainty of associated input model parameters. Hence, there may be subspaces worthy of scrutiny based on possible application-specific interactions utilizing algebraic sensitivity analysis, for example (Norton, 2008). In such cases, subspace explorations may be relatively straightforward. However, if obvious interactions are not immediately forthcoming from the mathematical model, it is often unproductive to seek relevant subspace interactions. Hence, augmenting standard RSA with methods sensitive to parameter interactions is of high priority, and we do so in the context of the dengue model.

### 4. Augmenting Standard RSA With MLMs

Our augmentation of standard RSA utilizes two MLMs based on decision trees that address all three aspects of global sensitivity analysis (GSA) that have been enumerated by Saltelli et al. (2008) and more recently by Pianosi et al. (2016), but in the context of RSA’s binary classification of “passing” or “failing” parameter vectors. These three purposes of GSA exercises are as follows:

#### Table 1

| Parameter | Description                                      | Typical values          |
|-----------|--------------------------------------------------|-------------------------|
| $\mu_{\text{E}}$ | Egg mortality rate (per day)                       | 0–0.1                   |
| $\theta$ | The ratio of minimum egg hatching rate to ideal egg hatching rate (dimensionless) | 0–1                     |
| $\lambda$ | The ratio of minimum larvae development rate to ideal larvae development rate (dimensionless) | 0–1                     |
| $\omega_0$ | The maximum heavy rain washout fraction (dimensionless) | 0–1                     |
| $\omega_{\text{max}}$ | Highest water level in the system, beyond which overflow will occur (mm) | 200–2000                |
| $\delta$ | Ratio between minimum and maximum water level (dimensionless) | 0–1                     |
| $\pi_{\text{max}}$ | Maximum carrying capacity for immature stages (mosquito) | $1.0 \times 10^5$–$1.2 \times 10^7$ |
| $\nu_{\text{em}}$ | Duration from emerging adults to adults (day) | 1–7                     |
| $\mu_{\text{em}}$ | Mortality during adult emergence (per day) | 0–0.2                   |
| $\sigma$ | The ratio of infected to uninfected immature and mature mosquito death rate (dimensionless) | 1–3                     |
| $\tau_{\text{exh}}$ | Intrinsic incubation period (day) | 3–9                     |
| $\tau_{\text{ih}}$ | Recovery period (day) | 3–9                     |
| $\omega_{\text{vh}}$ | Transmission probability from vector to human (dimensionless) | 0–1                     |
| $\omega_{\text{hv}}$ | Transmission probability from human to vector (dimensionless) | 0–1                     |
| $\varphi$ | Reporting rate (dimensionless) | 0–1                     |
| $\delta_{2013}$ | Timing for the imported case in 2013 (1 January 2012 as Day 1) | 521–571 (1 January 2012 as day 1) |
| $\delta_{2014}$ | Time for the imported case in 2014 (1 January 2012 as Day 1) | 853–903                 |
| $\mu_a$ | The survival rate for adults after intervention (per day) | 0–1                     |
| $\mu_i$ | The survival rate for immature stage after intervention (per day) | 0–1                     |

*Note. See references for the typical values of each parameter from Cheng et al., 2017.*
1. Ranking (or Factor Prioritization) aims to sort the vector of parameters (ξ) according to each parameter's relative contribution to output variance.

2. Screening (or Factor Fixing) aims to identify which, if any, parameters in ξ have a negligible influence on output variance.

3. Mapping, which aims to determine the regions of parameter space that produce output values of particular interest.

Use of MLMs to augment RSA is critical to distinguish ranking and screening in complex models, as noted by Pianosi et al. (2016). The Kolmogorov-Smirnov statistic, useful for ranking, cannot be used for screening because the univariate distribution of a parameter that is truly insensitive and one that only affects output through interactions may be indistinguishable; this fact motivates in part our interest in augmenting RSA.

Mapping remains based on the pass/fail criterion due to its simplicity and ability to integrate a wide variety of possible model output with various notions of what constitutes a pass: either a more traditional calibration criterion or a “forward” approach to guide control or data collection efforts as proposed in the earliest work on RSA (Spear, 1970; Spear & Hornberger, 1980). In addition, we address the issue of scale of sensitivity, an inherent aspect of GSA, but whose implications are seldom explored. In this context, scale refers to the phenomenon in which parameters regarded as unimportant based on ranking performed in the entire parameter space may nonetheless be highly influential in small subspaces, with implications for the screening aspect of GSA. The scale issue has recently been addressed directly by Haghnejahdar and Razavi (2017) and in the variogram-based SA approach of Razavi and Gupta (2016). Here, we address scale from the perspective of RSA, which uses Monte Carlo simulations to examine the effect of perturbation scale on pass probability. Also, using Monte Carlo simulation to average out effects from other parameters, we examine perturbation scale dependency of a focal parameter on passing probability. Additionally, we examine the effects of scale upon parameter screening generally by MLMs that are built in subspaces we identify as being potentially worth further investigation; both techniques are described below in context of the dengue model.

In the mid-90s, our group began to experiment with tree-structured MLMs to explore parameter interactions as well as mapping the passing space more deeply (Spear et al., 1994). Despite early promise, these methods were not widely applied in sensitivity analysis until quite recently (Aulia et al., 2019). The basis for most later development of tree structured methods was CART, Classification and Regression Trees (Breiman, 2001). When compared with statistical models, CART is a nonanalytic, computationally intensive procedure which partitions parameter space into high-dimensional “bins” such that a chosen partition minimizes an error (loss) function which depends on the application (regression or classification). In either case, CART works by creating trees based on binary splits of the data. Splits are phrased as binary questions; is ϕ ≤ c? where ϕ is a parameter in the set ξ and c a value in ϕ's allowed range. The split generates two new nodes: When a new parameter vector encounters the split, if the condition is met, it is allocated to the left node, and the right node otherwise. The goal at each split is to make each node as homogeneous as possible as measured by the error function. The set of binary splits leads to a partitioning of parameter space such that a new parameter vector can be uniquely assigned to a bin (Spear et al., 1991).

In practice, algorithms that construct such trees must consider at each step which ϕ ∈ ξ to split on and also at what point c to make the split. This is accomplished by exhaustive search, first over all ϕ ∈ ξ and then over all c ∈ range(ϕ). At each step in making the tree, the CART algorithm greedily selects the best (ϕ, c) combination that will improve the error function. In this case, “greedy” means that the algorithm is not allowed to “look back” and change splits closer to the root in order to improve error. Because if left unchecked, CART can improve error to arbitrarily low levels by continuously making binary splits until each data point is in a single node, a stopping rule must also be applied; typically, the splitting process continues until the number of data points in a node becomes smaller than some critical value (Spear et al., 1991).

Once a large tree is grown and the algorithm terminates, a postprocessing step known as pruning begins. The goal of pruning is to prevent overfitting, the phenomenon where a large, complex tree might have very low error on a “training” data set used to build it but do very poorly on a “testing” data set. Pruning proceeds by cutting branches in the tree according to a tuning parameter, usually denoted α, which governs the trade-off between simpler trees and performance on training data. To choose α, cross-validation is normally used, in which the data is randomly split into v subsets; v − 1 subsets are used to construct the tree, and the final subset is used to estimate error. By repeatedly splitting the data set randomly and evaluating error, an optimal
The value of $\alpha$ can be found that minimizes error on test data; the reduced complexity of the pruned tree also improves interpretability (James et al., 2013).

An example of a tree structure produced by the CART algorithm when applied to the dengue model is shown below in Figure 1. Because the analysis was concerned with what parameter subspaces lead to passing versus failing simulations (classification), the Gini impurity was used as the error function to construct the tree. Minimizing the Gini impurity results in nodes where all data points belong to a single class as much as possible. In the simple binary case, the Gini impurity has the form

$$IG = 1 - \left(p^2 + (1-p)^2\right),$$

where $p$ is the probability of class membership. The Gini impurity reaches a maximum value when $p = 0.5$ and decreases monotonically as $p$ approaches 0 or 1, due to the quadratic form of the function (Hastie et al., 2009). The error function depends only upon the proportions of passing versus failing simulations in a single node, where it is minimized by exhaustive search.

Examination of the resulting tree reveals $\omega_{\text{max}}$, which defines the maximum water availability, to be the most important single parameter in initially discriminating passing from failing parameter vectors. Although extensive experience with CART often produced interpretable results, it was found to have a tendency to overfit training data despite pruning. That is, a small change in the input data could result in a large

**Figure 1.** An example CART diagram generated from analysis of the dengue model. Internal nodes are represented as binary questions, and terminal nodes in the tree correspond to the high-dimensional (five-dimensional, in this case) bins into which new parameter vectors can be allocated. The first row in each box represents the category this node is classified as, according to the majority class in that node (e.g., if the proportion of passing parameter set values is greater than 0.5, then that node is classified as “Pass”). The two numbers in the second row represents the proportion of fails and passes in this node, respectively. The third row represents the percentage of all sets of parameter values in this node. As the tree goes deeper, the purity of each node increases.
change in the structure of the estimated tree. Combatting overfitting motivated the subsequent development of various elaborations, notably Leo Breiman’s widely used random forest algorithm (Breiman, 2001), which is applied here for parameter ranking and screening of the dengue model.

Random forest is based on the observation that overfitting in decision trees is often due to high variance between individual trees, meaning that if the same estimation procedure is applied to new data sampled from the same population, the resulting tree may look quite different. However, if many trees are individually estimated and then the predictions are averaged, the resulting ensemble is much more stable (lower variance) than a single tree. To construct the ensemble, select some large number $B$ of trees; for each tree, we (bootstrap) resample the data set to provide training data; this procedure is known as bagging.

Random forest goes one step further: When building each tree in the forest, the splitting algorithm may not choose from the entire vector $\xi$ of parameters to find the optimal split, but some randomly sampled subset of parameters. This additional random sampling reduces correlation between individual trees, which reduces the variance of the overall prediction. If used for classification, the final ensemble prediction from a fitted random forest model for each data point uses a majority vote rule from the individual trees. This strategy is reported to perform very well against competing approaches (Liaw & Wiener, 2002). We use the R packages randomForest to implement this analysis (Liaw & Wiener, 2002). While this approach helps address overfitting, it is at the cost of increased complexity of interpretation and visualization of the results as will be seen.

Finally, tree-structured methods have been applied to the task of multivariate density estimation, and in the RSA context, mapping the density of parameter vectors resulting in passes within the parameter space. Early in our explorations, we applied a tree-structured density estimation (TSDE) technique to map the pass space (Spear et al., 1994). TSDE modifies the CART concept for the task of density estimation: Given a set of passing parameter vectors, how might we estimate the probability density function that gave rise to those passes? From the point of view of CART, modification of the algorithm revolves around finding a suitable error function. In TSDE, the integrated squared error (ISE) was used, the form of which is given below, where $\hat{f}$ is the estimated density:

$$ISE = \int_{x \in S} \left( f(x) - \hat{f}(x) \right)^2 \, dx,$$

Where $S$ is the total parameter space (the space of $\xi$). Because our method requires that each $\varphi \in \xi$ have some range given by minimum and maximum bounds, $S$ will be a hyperrectangle whose dimension is the same as the number of parameters (size of $\xi$). Although the formal definition of ISE relies on the unknown density $f$, a simple plug-in estimator based only on the estimated function can be used. We consider further properties of TSDE in the context of the dengue example.

5. Parameter Ranking Based on Random Forests

As noted previously, standard RSA using the Kolmogorov-Smirnov statistic will not effectively deal with parameter interactions, making parameter screening based on it unreliable. The use of decision trees generally, and random forests in particular, allows greater confidence in parameter importance-based ranking because interactions have been incorporated in the analysis. For example, the $x$-axis of Figure 2a is a measure of the mean decrease in impurity associated with each parameter as measured by the Gini impurity defined above, and averaged over all splits where this parameter was selected and all trees in which it was selected during the process. Larger values indicate more important parameters (Hastie et al., 2009). Because a parameter can be selected multiple times on a single branch of a tree or on various branches, the procedure can detect effects and interactions at varying scales.

The importance ranking shown in Figure 2a results from analysis of the Cycle 2 dengue data using 500 trees (in that context, the issue arises of how many trees are needed to produce stable results, an issue addressed directly by Behnamian et al. (2017)). Unlike the CART tree shown above, however, because multiple trees are involved, there is no single tree that can be used to summarize the results for an ensemble of trees except in some average sense (Chipman et al., 1998).
The random forest importance ranking based on the Cycle 2 data was generally consistent with the parametric difference ranking based on \(d_{m,n}\) as shown by the scatter plot in Figure 2b, but there are two notable exceptions. First, the parameter \(\alpha_{vh}\) shows a larger importance in the random forest ranking than in the \(d_{m,n}\) ranking. Similarly, \(\mu_{a}\) shows a large \(d_{m,n}\) versus Gini differential consistent with the interactions with \(\mu_{i}\) as will be seen in Figures 3a and 3c below.

As shown in Figure 2b, at the scale of the Cycle 2 analysis, four out of five parameters that resulted from \(d_{m,n}\) were selected as the least important by the random forest analysis. However, the random forest analysis allows greater confidence to be placed in the ranking of importance because interactions and local effects within the parameter space have also been explored. As noted above, it is not generally straightforward to identify these interactions and estimate their importance. This challenge motivated the recent development of partial dependence plots (Friedman, 2001) and forest floor visualization algorithm (Welling et al., 2016) that provides tools to address these issues that are particularly challenging in the majority of applications.

Figure 2. Variable importance. (a) Variable importance measured by Gini impurity and (b) the scatter plot of variable importance measured by Gini impurity and univariate marginal Kolmogorov statistic \(d_{m,n}\) using deterministic Cycle 2 dengue data.

The random forest importance ranking based on the Cycle 2 data was generally consistent with the parametric difference ranking based on \(d_{m,n}\) as shown by the scatter plot in Figure 2b, but there are two notable exceptions. First, the parameter \(\alpha_{vh}\) shows a larger importance in the random forest ranking than in the \(d_{m,n}\) ranking. Similarly, \(\mu_{a}\) shows a large \(d_{m,n}\) versus Gini differential consistent with the interactions with \(\mu_{i}\) as will be seen in Figures 3a and 3c below.

As shown in Figure 2b, at the scale of the Cycle 2 analysis, four out of five parameters that resulted from \(d_{m,n}\) were selected as the least important by the random forest analysis. However, the random forest analysis allows greater confidence to be placed in the ranking of importance because interactions and local effects within the parameter space have also been explored. As noted above, it is not generally straightforward to identify these interactions and estimate their importance. This challenge motivated the recent development of partial dependence plots (Friedman, 2001) and forest floor visualization algorithm (Welling et al., 2016) that provides tools to address these issues that are particularly challenging in the majority of applications.

Figure 3. Different visualizations of the interaction between \(\mu_{a}\) and \(\mu_{i}\). (a) Partial dependence plot; (b) feature contribution plot for \(\mu_{i}\) from forest floor; and (c) feature contribution plot for \(\mu_{a}\). Colors in (a) represent the predicted passing probability. In (b) and (c), the x axis represents the values for \(\mu_{i}\) and \(\mu_{a}\), respectively, and y axis shows their feature contributions.
of random forests. These applications analyze statistical data that do not arise from the type of structured mathematical models that RSA was developed to address. These two methods were implemented with R packages pdp (Greenwell, 2017) and forestFloor (Welling et al., 2016), respectively.

The partial dependence plot shows the relationship between parameters of particular interest and the predicted passing probability, which is the mean probability of model output meeting the “pass” criteria across all values of all other parameters (Molnar, 2019). Interactions can be shown when the partial dependence of two parameters are plotted at once. Figure 3a shows the two-dimension partial dependence plot between $\mu_a$ and $\mu_4$. When $\mu_4$ increases, the influence of $\mu_a$ on passing probability also changes. Forest floor visualizes out-of-bag feature contributions, which are the total local increments across all nodes split by a particular parameter and all trees standardized by the number of trees in the out-of-bag samples. For classification, which is of interest here, the local increment refers to the increase in passing probability from parent node to daughter node. In forest floor visualizations, interactions can be reflected by two means, the goodness-of-visualizaton (GOV, noted as $R^2$ in Figures 3b and 3c) and the pattern of the color gradient. GOV is the square of the Pearson’s correlation coefficient between estimated feature contributions from a one-way feature contribution plot and the observed feature contributions. Higher GOV means more variance has been explained by the main effect, suggesting remaining interactions involving the focal parameter are either absent or relatively unimportant. In Figures 3b and 3c, the GOV are 0.79 and 0.59, respectively, indicating that weak interactions might be present. On the other hand, if the feature contribution plot of parameter $\xi_i$ was colored by the value of another parameter $\xi_j$ in the same parameter vector, for example, with red and blue representing small and large $\xi_i$, respectively, the occurrence of a vertical color gradient in the plot suggests an interaction between $\xi_j$ and $\xi_i$, because similar values of $\xi_j$ led to different impacts of $\xi_i$ on the predicted probability of pass. For example, when colored by the value of $\mu_i$ (Figure 3b), the feature contribution plot of $\mu_4$ shows a vertical red-to-blue gradient (Figure 3c), indicating the interaction between these two parameters. The full set of one-way feature contributions for all parameters colored by $\mu_i$ are shown in Figure S7.

Prior to addressing the application of RSA and random forests to ranking parameter importance in the stochastic dengue model, we briefly address the possibility of using the foregoing results to provide more targeted exploration of the parameter space to increase pass rates, especially for computationally expensive simulation runs. Again, from the forest grown from an equal number of simulated passing and failing parameter vectors comprising the training set described above, recall that each is classified as a pass or fail by each tree of the ensemble. Figure S8 shows the distribution of the probability of each parameter vector being predicted as pass, which is calculated as the proportion of trees predicting it as a pass, grouped by simulation results. As can be seen, there are simulated passes being correctly classified by as few as 25% of trees and true failing vectors misclassified as passes by over 75% of trees. However, in the latter case, there is a relatively small likelihood of misclassification of a pass if we define that being predicted to be a pass be defined as one in which a proportion of 75% or greater of trees classify it as such. Hence, suppose a second cycle is conducted in which parameter vectors are randomly selected from the prior distributions and each is run down each tree of the ensemble. Then, only those vectors being classified as a pass by 75% or more trees, for example, are selected for simulation. Table S2 shows the results of exploring this strategy using runs of about 50,000 simulations when 75% or 50% was used as the threshold for classifying passes.

Only 66 of 270 simulated passes were predicted to be passes with a probability of 0.75 or greater by the ensemble of trees. While following this option would increase the pass rate to 4.2% from the initial 0.54%, 204 passes or 75% of all simulated passes were misclassified. As might be expected, reducing the pass probability cutoff to 0.50 results in a higher fraction of simulated passes being correctly classified, 226 of 270, but increases the number of incorrectly predicted passes to 10621 for a pass rate of 2.1%.

### 6. Stochastic Models

We developed a stochastic version of the dengue model by interpreting the differential equations as describing a discrete-event process, using the same parameters and functional relationships as in the deterministic analysis (Cheng et al., 2017). However, instead of simulating each state variable according to the differential equations directly, we now simulate the occurrence of each event, such as the death of a mosquito, the recovery of a human case, and the infection of a mosquito, and then aggregate these events to get the state
Table 2

| Number of additional passes | Initial class |
|-----------------------------|---------------|
|                             | 0  | 1  | 2  | 3  | 4  |
| Fail                        | 493| 7  | 0  | 0  | 0  |
| Pass                        | 226| 99 | 41 | 14 | 1  |

variable. Since events can occur at different frequency and order, the simulated time series of new cases and the corresponding pass/fail classification of the same parameter values can change from run to run.

Early in its history, it was recognized that the RSA procedure could be applied to stochastic models (Beven & Binley, 1992). However, stochasticity poses additional challenges to the search for adequate numbers of passing parameter vectors to support subsequent analyses. The fundamental distinction between the deterministic and stochastic models is the unique assignment of a simulation as a pass or fail in the deterministic case does not carry over to the stochastic case. In the latter, the analog is the probability of a pass or a fail for any parameter vector. In the analysis of the deterministic dengue model, it became clear that the most likely causes of the 2014 outbreak were the timing of the first imported case which led to local transmission in 2013 and 2014 and the rainfall in 2014. Earlier timing of local transmission could be explained by earlier imported cases, but there was also the possibility that pure chance was important for a single imported case. For example, if a mosquito bites an infected person, does it then survive to infect others? The investigation of these competing explanations of the outbreak was the objective of employing a stochastic model. Hence, the first question that arose in applying RSA to the stochastic case was the differential importance of parametric uncertainty and stochastic variability.

Using the same priors on the fixed parameters of the model used in the Cycle 2 deterministic analysis reported above led to an initial set of 381 passing simulations of the stochastic model among 200,000 simulations. In order to assess the impact of stochastic variation versus parametric uncertainty, we added 500 randomly chosen vectors from failing simulations and then conducted 12 further simulations at each of those 881 points in parameter space. Table 2 shows the number of additional passes observed about each point. Surprisingly, only 155 of the 381 passing parameter vectors led to one or more further passes about the same fixed parameter vector. If we regard a pass point with no further passes in the 12 trials as a "stochastic pass" and a pass with 1 or more further passes as a "parametric pass," this results in an estimate of 59% of the total passing simulations as stochastic passes. This observation suggests that, at least in the regions of parameter space sampled here, fundamental variance in trajectories from the stochastic model played an important role in the observed epidemic in Guangzhou. Examination of the parameter vectors initially classified as fails showed nearly all (493) to be parametric fails, that is, stochastic variance was not sufficient to overcome unsuitable epidemic conditions. Clearly, conditions had to be right for an outbreak to occur, but when they were, bad luck played a big role.

Accumulating the 381 passes used in the foregoing analysis required 200,000 total simulations for a pass rate of 0.19% compared to the 0.54% found in the deterministic analysis with the same prior parameter distributions. Figure 4 shows the variable importance plot resulting from the 381-pass training set and 500 fails. This is the analog to Figure 2a for the deterministic analysis. As can be seen, the variable importance scores are substantially less in the stochastic case with a maximum score of about 40 versus a value of slightly over 100 in the deterministic case, presumably reflecting the "noise" introduced by stochastic effects. In addition, the order of importance is different between the stochastic and deterministic cases although eight of the 10 most important parameters are the same in the two cases. The most notable differences in the order of importance are $\omega_{\text{max}}$ changing from first in the deterministic case to sixth in the stochastic and $\alpha_{\text{ch}}$ changing from a deterministic seventh to first place in the stochastic ranking.

### 7. Mapping and Equifinality

Recall that mapping aims at determining the regions of the parameter space that produce output values of particular interest. In the RSA context, the pass criteria often define "particular interest" as those simulations consistent with the observed behavior of the system being modeled. Alternatively, in applications involving system design or modification, the pass criteria may define the desired behavior of the system. In both cases, there will be many parameter vectors which meet the pass criteria rather than a single optimal vector as noted earlier. This is generally true in any application of RSA and exemplifies the concept of equifinality. As in the dengue example, there may be few passes in the tails of some univariate priors which motivate trimming to increase the pass rate in a second run. However, any such reduction of the parameter space...
will exclude some number of simulations which led to outcomes consistent with the prior definition of acceptable. Barring any redefinition of acceptability or new information on the prior distributions, focusing on any particular parameter subspace carries with it the question of the probability of the “real” system being best modeled by parameter values in that subspace. This becomes a more urgent question when passes in different parts of the pass space lead to different explanations of the factors controlling system behavior or different intervention strategies. While this is largely an application-specific issue, mapping is important in identifying such possibilities. From a statistical perspective, this is a problem in density estimation.

Early in our explorations of tree-structured methods, we applied a TSDE technique to map the pass space (Spear et al., 1994). TSDE modifies the CART concept for the task of density estimation: Given a set of passing parameter vectors, how might we estimate the probability density function that gave rise to those passes?

Here, we applied a more recent implementation of that approach to the Cycle 2 dengue data, density estimation trees (DETs), which is better than other density estimators in two specific forms of model adaptability: adaptability between dimensions (ABD) and adaptability within a dimension (AWD) (Ram & Gray, 2011). Because their treatment regards DET as a nonparametric density estimator, ABD is concerned with behavior of the estimator when unimportant dimensions are added; from a statistical point of view, if random noise is added as an additional dimension, DET should “smooth” away that unimportant dimension. AWD concerns the ability of the estimator to adapt to local behavior, essentially, to pick up on interesting regions and avoid “oversmoothing” genuine structure. These two features of the method correspond to two of the purposes of GSA we are concerned with, namely, screening and mapping. The method is adapted from the C++ machine learning library MLPACK (Curtin et al., 2013) to R (https://github.com/slwu89/density-estimation).

Figure 5 shows the resulting tree. The distribution of each parameter was converted to a standard uniform distribution through probability integral transformation before running the DET algorithm. The value within each of the nodes is the density of that subspace relative to the overall density in the entire space. Six of the seven parameters shown in the figure are at or near the top of the list of importance identified by the random forest analysis.

The two main branches of the tree split initially on \( \omega_{\text{max}} \), which defines the highest annual water level in the system. Hence, the split divides between scenarios having high and low water availability with somewhat different parametric determinants of an outbreak in the two cases. When \( \omega_{\text{max}} \) is low, which represents limited environmental carrying capacity for mosquitoes, the observed timing and magnitude of the dengue epidemics was likely to be mainly controlled by the timing of imported cases (\( \beta_{2013} \) and \( \beta_{2014} \), see the definitions of each parameter in Table 1); otherwise, they were mainly determined by the effectiveness of the interventions (\( \mu_a \) and \( \mu_i \)). We chose to further explore a high-density node from each branch by a directed search targeted at the 4.41 density node in low water availability scenario and the 3.32 density node in high water availability scenario (highlighted nodes in Figure 5). The probability of a random sample from the Cycle 2 space lying within those subspaces is about 0.27 and 0.14, respectively.

The subspace trees are shown in Figures S9 and S10. Both the high and low maximum water availability nodes led to more detailed trees with new parameters appearing, \( \alpha_{\text{vh}} \), \( \alpha_{\text{hv}} \), and \( \pi_{\text{max}} \), in addition to those appearing earlier. The highest density nodes in these subspace trees relative to the density in the original space are 11.3 (3.32 × 3.4) and 14.1 (4.41 × 3.19) discounting the edge effect of \( \mu_{a} \) at the bottom of the latter node. Recalling, however, that the pass rate in the Cycle 2 space was about 0.5%, the pass rate in these two subspaces has increased by an order of magnitude but is still relatively low at 6% or 7%. This implies that the volume of the pass space within the subspaces of the DET is still low in this case, as it was in the earlier to the
MMSOILS model (Spear et al., 1994). As noted above, this observation also suggests yet more complex structure at smaller scales, which led us to investigate the lower limit on scale, that is, local sensitivity about pass points.

Fifty of the Cycle 2 pass points were randomly chosen from the original 948. As before, all the parameters are transformed to standard uniform distributions before the analysis. Each individual parameter was then perturbed in ±0.01 increments and the passing fraction among the 50 random pass points was determined in model runs. This fraction at each perturbation was calculated about each of the 19 parameters and plotted as shown in Figure 6. If a perturbation lies beyond the boundary of a particular parameter, it was set at that boundary value. Hence, if that boundary value is a pass or a fail for a particular pass point, all greater perturbations of that parameter beyond the boundary will be so as well. As can be seen, for negative perturbations of $\omega_{\text{max}}$, boundary values led to a flatterening of the curve.

The most striking features of Figure 6 are the narrow and overlapping curves for $\pi_{\text{max}}$, $\mu_{\text{em}}$, $\alpha_{\text{vh}}$, and $\alpha_{\text{hv}}$, suggestive of significant interactions among them. If we regard the $y$-axis, the pass status of perturbations in each parameter direction, as points on a response surface, the $\pi_{\text{max}}$, $\mu_{\text{em}}$, $\alpha_{\text{vh}}$, and $\alpha_{\text{hv}}$ patterns suggest a narrow ridge on that surface. Despite the fact that the marginal distributions of these parameters indicate that passes occur over nearly their entire range of values and the importance ranking placed these parameters in the low range to mid-range at the scale of the Cycle 2 space, in the neighborhood of pass points, these are the most sensitive parameters. We replicated this analysis with a second set of 50 randomly chosen parameter vectors as well as constructed the equivalent plots for the high and low water availability subspaces. Some modest differences were seen, but the overall patterns are very similar. These patterns suggest a multidimensional rope winding from corner to corner of the 19-dimensional space but whose cross section is quite narrow in these four dimensions. These local patterns provide one explanation of why the pass rates remain modest even in the higher density nodes of the tree.

We conducted several runs to assess the impact of interactions in the foregoing four-dimensional subspace on the overall pass rate by initially fixing the remaining 15 parameters plus $\mu_{\text{em}}$ at 100 randomly selected

Figure 5. Density estimation tree for deterministic Cycle 2.
pass values from Cycle 2, then sampling from the three-dimensional kernel density conditioned on $\mu_{em}$ to set the other three parameter values. Kernel densities were estimated with the R package \textit{ks} (Duong, 2007). The overall pass rate was 8.76%. Following the same procedure for each of the other three in turn resulted in pass rates from 8.19% to 8.82%. These values are in the same general range as in the highest density nodes in the tree shown in Figure 5 but arise from vectors randomly drawn from 16 of the 19 dimensions of the entire posterior space. This may suggest a two-pronged approach to characterizing the posterior space that utilizes information from both tree structured analyses and subspace characteristics. Identification of subspaces via tree-structured methods to be followed up with more intensive statistical methods is in the spirit of current research (Gramacy & Lee, 2008) and may offer further insight into the structure of the parameter distribution.

8. Screening and Scale

We have left the issue of screening until last since the importance of scale is so clearly illustrated in this application of RSA by the mapping results obtained with the DETs supplemented by the local sensitivity results. Another example is from the DETs of the Cycle 2 space, Figure 5, versus the subspace trees at the high and low water availability nodes, Figures S5 and S6. Only $\mu_{i}$ and $\mu_{a}$ occur in all three trees and $\omega_{\text{max}}$ only in the full space tree, although there it is ranked first in importance. Hence, screening is largely an issue of deciding on the scale of sensitivity. Interestingly, five parameters originally mentioned as candidates for screening on the basis of $d_{m,n}$, $\theta$, $\sigma$, $\tau_{\text{exh}}$, $\tau_{\text{inh}}$, and $\delta$ do not appear further at the scale of the high/low water availability subspace nodes although their local sensitivity shows low to medium sensitivity among them. In general, however, the importance of screening seems to be highly application specific. From a methodological perspective, however, others have made the case for the importance of scale on sensitivity analysis generally, and methods have been proposed to quantify it (Haghnejahdar & Razavi, 2017).

9. Discussion

Recall that our central objective was to determine if modern statistical methods utilizing machine learning might improve the usefulness of RSA while preserving the conceptual simplicity of its original formulation. Our experience in the dengue studies suggests that the use of random forests and DETs are both useful additions to the original RSA approach. In this application, we found that the original RSA results based on the marginal distributions of passing versus failing parameter vectors using the Kolmogorov statistic remain useful. However, random forest models can directly assess the effects of parameter interaction on importance rankings. The random forest importance ranking was generally consistent with the parametric

![Figure 6. Local sensitivity plot. The percentage of passes among 50 randomly sampled passes after perturbation.](https://example.com/figure6.png)
difference ranking based on $d_{m,n}$, as shown by the scatter plot in Figure 2b, but there are two notable exceptions. First, the parameter $\mu_x$ shows a large $d_{m,n}$ versus Gini differential consistent with the interactions with $\mu_y$ seen strikingly in Figures 3b and 3c. Also, $\omega_{xc}$ shows a larger importance in the random forest ranking than in the $d_{m,n}$ ranking. The Forest Floor software provides a means of exploring these interactions further and also may provide a link between the random forest results and the density estimation results as noted below. While forest floor has been published recently, random forests have been applied to statistical data for some years and a substantial literature exists. However, the RSA application to mechanistic mathematical models is rather different. Experience with both will hopefully reveal variations on the theme particularly useful for those applications.

A notable result of the application of RSA to the stochastic dengue model was the exploration of the relative importance of stochastic variation compared to parameter uncertainty encoded by the prior distribution on pass probability. Our particular means of estimating this probability was clearly quite simple, and more powerful approaches are no doubt possible. However, the estimation of the effects of chance versus parameter uncertainty can be considered an important form of macrosensitivity of stochastic models. In the context of the dengue model application, the identification of conditions ripe for an outbreak of disease is of general importance in epidemiological studies and, presumably, in many other situations where weather conditions, for example, exert a strong influence on the likelihood of outcomes of interest or concern.

The use of DETs we believe to be particularly useful as a means of emphasizing the importance of equifinality. In many applications of mathematical modeling under parametric uncertainty, rigorous exploration of the implications of parameter variability has on the relative support given to different causal explanations of system behavior (equifinality) is crucial for understanding and planning of interventions. In addition, DETs illustrate the fact that sensitivity depends on scale. Different parameters emerge as important as smaller subspaces lower in the tree are explored. Hence, the method underscores that scale is important to the screening objective of sensitivity analysis methods. Fixing any subset of parameters implies that their influence on conclusions drawn from subsequent analyses is not of practical importance. However, in other applications, the local sensitivity results presented above are cautionary.

We have only mentioned in passing an often-noted disadvantage of the RSA approach, the difficulty of finding enough passing simulations to support the use of some of the methods we have been able to investigate in the dengue example. By focusing on interesting subspaces and targeted sampling, we were able to increase the pass rate by about an order of magnitude. Nevertheless, we achieved pass rates of less than 10% at best. Emulators have been used to identify particular subspaces for further analysis (Yang et al., 2018), and others have approached related issues from the local sensitivity perspective (Jefferson et al., 2016). However, if the initial number of passing simulations is adequate to support a DET analysis, then the subspaces of particular interest for future analysis can be selected and, importantly, regions of the pass space nonetheless excised from subsequent study are clearly defined.

The adequacy of the number of passing simulations to support a DET analysis raises another methodological issue, that of overfitting a data set with a single tree, which is the issue that afflicted CART trees and that random forests was developed to solve. As noted by Ram and Gray (2011), there is some tension between the goals of adaptability, defined as both adaptability between and within dimensions (ABD and AWD) and interpretability, that is, answering questions about the density related to variable/dimension importance and simple univariate splitting rules. Despite this tension, their sophisticated implementation of DET was shown in the paper in which it was introduced to have cross-validated error comparable to local kernel density estimators. The algorithmic methods used to combat overfitting follow those of Breiman et al. (1984), modified to work with a loss function appropriate for the task of density estimation. An optimal regularization parameter is selected through cross-validation to prune the tree. Considering the difficulty in managing this tension, we believe that DET is the best modern tree-based density estimation tool in a high-quality open source software package (Curtin et al., 2013). While overfitting resulting in instability remains an issue, these methods are unbiased with respect to density estimation (Anderson & Burnham, 2004). Furthermore, while far from a panacea, especially when model runs are computationally expensive, further simulation runs allows additional samples to be drawn from the target density. Happily, the RSA concept is not tied to any one algorithm, and current research in this vein may prove useful to the analysis of mechanistic models (Meyer, 2018).
In summary, we suggest that, insofar as there is a standard RSA, it includes both a random forest analysis to incorporate interactions in parameter importance ranking and the use of DETs for mapping and to inform screening. Also, in the mapping context, we encourage consideration of the implications of application-specific equifinality as a common element of RSA applications. Overall, we believe that our explorations of machine learning techniques developed over the last several decades offer significant but easily implemented improvements in the analysis of mechanistic models that RSA was originally developed to address.

References

Anderson, D., & Burnham, K. (2004). Model selection and multi-model inference. (Second ed.p. 63). NY: Springer-Verlag.
Aulia, A., Jeong, D., Saad, I. M., Kania, D., Shuker, M. T., & El-Khatib, N. A. (2019). A random forests-based sensitivity analysis framework for assisted history matching. Journal of Petroleum Science and Engineering, 181, 106237. https://doi.org/10.1016/j.petrol.2019.106237
Beck, M. B. (1987). Water quality modeling: A review of the analysis of uncertainty. Water Resources Research, 23(8), 1393–1442.
Behnamian, A., Millard, K., Banks, S. N., White, L., Richardson, M., & Pusher, J. (2017). A systematic approach for variable selection with random forests. Achieving stable variable importance values. IEEE Geoscience and Remote Sensing Letters, 14(11), 1908–1922.
Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes, 6(3), 279–298.
Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology, 249(1-4), 11–29.
Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.
Breiman, L., Friedman, J., Stone, C., & Olshen, R. (1984). Classification and regression trees. Boca Raton: CRC Press.
Cheng, Q., Jing, Q., Spear, R. C., Marshall, J. M., Yang, Z., & Gong, P. (2017). The interplay of climate, intervention and imported cases as determinants of the 2014 dengue outbreak in Guangzhou. PLoS Neglected Tropical Diseases, 11(6), e0005701. https://doi.org/10.1371/journal.pntd.0005701
Chipman, H. A., George, E. I., & McCulloch, R. E. (1998). Extracting representative tree models from a forest.
Curtin, R. R., Cline, J. R., Slagle, N. P., March, W. B., Ram, P., Mehta, N. A., & Gray, A. G. (2013). MLPACK: A scalable C++ machine learning library. Journal of Machine Learning Research, 14(1), 801–805.
Duong, T. (2007). ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. Journal of Statistical Software, 2(7), 1–16.
Fenwick, D., Scheidt, C., & Caers, J. (2014). Quantifying asymmetric parameter interactions in sensitivity analysis: Application to reservoir modeling. Mathematical Geosciences, 46(4), 493–511.
Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 1189–1232.
Gramacy, R. B., & Lee, H. K. H. (2008). Bayesian treed Gaussian process models with an application to computer modeling. Journal of the American Statistical Association, 103(483), 1119–1130.
Greenwell, B. M. (2017). pdp: An R package for constructing partial dependence plots. The R Journal, 9(1), 421–436.
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
Haghnejadah, A., & Razavi, S. (2017). Insights into sensitivity analysis of Earth and environmental systems models: On the impact of parameter perturbation scale. Environmental Modelling & Software, 95, 115–131.
Hamby, D. (1994). A review of techniques for parameter sensitivity analysis of environmental models. Environmental Monitoring and Assessment, 15(2), 135–154. https://doi.org/10.1007/BF00547132
Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning: Data mining, inference, and prediction, (2nd ed.). New York, NY: Springer.
Hornberger, G., & Spear, R. (1980). Eutrophication in Peel Inlet—I. The problem-defining behavior and a mathematical model for the phosphorus scenario. Water Resources, 14(1), 29–42.
James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning: With applications in R (Vol. 112). New York: Springer.
Japkowicz, N. (2000). The class imbalance problem: Significance and strategies. Paper presented at the Proc. of the Int'l Conf. on Artificial Intelligence.
Jefferson, J., Gilber, J., Constantine, P., & Maxwell, R. (2016). Active subspaces for sensitivity analysis and dimension reduction of an integrated hydrologic model. Computers & Geosciences, 90, 78–89.
Llau, A., & Wiener, M. (2002). Classification and regression by randomForest. R news, 2(3), 18–22.
Meyer, D. W. (2018). Density estimation with distribution element trees.
Norton, J. P. (2008). Algebraic sensitivity analysis of environmental models. Environmental Modelling & Software, 23(8), 963–972.
Park, J., Yang, G., Satija, A., Scheidt, C., & Caers, J. (2016). DGSA: A Matlab toolbox for distance-based generalized sensitivity analysis of geoscientific computer experiments. Computers & Geosciences, 97, 15–29.
Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. Environmental Modelling & Software, 79, 214–232.
Poole, D., & Raftery, A. E. (2000). Inference for deterministic simulation models: The Bayesian melding approach. Journal of the American Statistical Association, 95(452), 1244–1255.
Ram, P., & Gray, A. G. (2011). Density estimation trees. Paper presented at the Acm Sigkdd International Conference on Knowledge Discovery & Data Mining.
Razavi, S., & Gupta, H. V. (2016). A new framework for comprehensive, robust, and efficient global sensitivity analysis: I. Theory. Water Resources Research, 52, 423–439. https://doi.org/10.1002/2015WR017558.
Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., et al. (2008). Global sensitivity analysis: The primer. Hoboken, NJ: John Wiley & Sons.
Sharefkin, M. (1983). Uncertainty and forecasting of water quality: Reflections of an ignorant Bayesian. In Uncertainty and Forecasting of Water Quality (pp. 373–379). Berlin, Heidelberg: Springer.

Spear, R. C. (1970). Monte Carlo method for component sizing. Journal of Spacecraft and Rockets, 7(9), 1127–1129.

Spear, R. C., Bois, F. Y., Woodruff, T., Auslander, D., Parker, J., & Selvin, S. (1991). Modeling benzene pharmacokinetics across three sets of animal data: Parametric sensitivity and risk implications. Risk Analysis, 11(4), 641–654. https://doi.org/10.1111/j.1539-6924.1991.tb00653.x

Spear, R. C., Grieb, T. M., & Shang, N. (1994). Parameter uncertainty and interaction in complex environmental models. Water Resources Research, 30(11), 3159–3169.

Spear, R. C., & Hornberger, G. (1980). Eutrophication in peel inlet—II. Identification of critical uncertainties via generalized sensitivity analysis. Water Research, 14(1), 43–49.

Teweldebrhan, A. T., Burkhart, J. F., Schuler, T. V., & Hjorth-Jensen, M. (2019). Coupled machine learning and the limits of acceptability approach applied in parameter identification for a distributed hydrological model. Hydrology and Earth System Sciences Discussion.

Welling, S. H., Refsgaard, H. H., Brockhoff, P. B., & Clemmensen, L. H. (2016). Forest floor visualizations of random forests. arXiv preprint arXiv:1605.09196.

Yang, J., Jakeman, A., Fang, G., & Chen, X. (2018). Uncertainty analysis of a semi-distributed hydrologic model based on a Gaussian process emulator. Environmental Modelling & Software, 101, 289–300.