Variance-Based Sensitivity Analysis: An illustration on the Lorenz '63 Model

Caren Marzban[1,2*]

[1] Applied Physics Laboratory, University of Washington, Seattle, WA 98195

[2] Department of Statistics, University of Washington, Seattle, WA 98195,

*Corresponding Author: marzban@stat.washington.edu

Abstract

Sensitivity Analysis (SA) generally refers to an assessment of the sensitivity of the output(s) of some complex model with respect to changes in the input(s). Examples of inputs or outputs include initial state variables, parameters of a numerical model, or state variables at some future time. SA is useful for data assimilation, model tuning, calibration, and dimensionality reduction; and there exists a wide range of SA techniques for each. This paper discusses one special class of SA techniques, referred to as variance-based. As a first step in demonstrating the utility of the method in understanding the relationship between forecasts and parameters of complex numerical models, here the method is applied to the Lorenz '63 model, and the results are compared with an adjoint-based approach to SA. The method has three major components: 1) analysis of variance, 2) emulation of computer data, and 3) experimental/sampling design. The role of these three topics in variance-based SA is addressed in generality. More specifically, the application to the Lorenz '63 model suggests that the Z state variable is most sensitive to the b and r parameters, and is mostly unaffected by the s parameter. There is also evidence for an interaction between the r and b parameters. It is shown that these conclusions are true for both simple random sampling and latin hypercube sampling, although the latter leads to slightly more precise estimates for some of the sensitivity measures.

1 Introduction

In contemporary times it is commonplace to represent complex systems with numerical models. Examples include numerical weather prediction models (Richardson 2007), ocean circulation models (Miller 2007), and hydrology models (Rushton 2003). All of these models generally consist of a system of partial differential equations which are numerically integrated, subject to boundary and initial conditions. Generally, such complex models can be viewed as a "black box" with some number of inputs and outputs. Although the choice of the inputs and outputs depends on the specific problem at hand, this paper focuses on model parameters and forecast quantities, respectively. Numerical models often have a large number of parameters whose values are not unambiguously known or even knowable, and so, it is useful to know how the parameters effect the forecasts.

Sensitivity Analysis (SA) is the name broadly associated with such questions, although it is performed for a variety of reasons, including variable/input selection, dimensionality reduction, data assimilation, and model tuning or calibration (Cacuci 2003). These different applications of SA are not necessarily mutually exclusive, but this paper focuses on the former. Specifically, the focus here is on the extent to which the outputs are affected by the various inputs across the full range of input values. In statistics "The full range of input values" is called the experimental region, and the question of how to choose it is a topic of text books on experimental design (Douglas 2005). It is important in SA because the sensitivity of the outputs to the

inputs can depend on the experimental region. It is unfeasible to examine the full experimental region, and so, sampling that region is usually the practical alternative. Two sampling schemes commonly employed in SA are simple random sampling and systematic sampling, reviewed below.

Given the long history of SA, there exists a wide range of methods for implementing it. One of the more intuitive is referred to as the "One-at-A-Time" (OAT) method (Saltelli et al. 2008, 2010). Generally, in an OAT analysis the inputs are varied one at a time, while all other inputs are held fixed at some value (e.g., at their respective mean), and the change in the output is monitored. However, if the number of inputs is large, the basic OAT approach samples only a small portion of the experimental region. This can be seen as follows: Consider three inputs whose values vary along the x, y, z Cartesian coordinates. Varying them one-at-a-time, will sample the points along the axes, but not at the corners of a cube centered at the origin. In three dimensions this is not a serious concern because one often assumes that the "black box" model is some relatively smooth surface, so that knowledge of its output values along the three axes is sufficient to define it uniquely. However, it is a geometric fact that a high-dimensional space consists mostly of corners (Jimenez and Landgrebe 1998; Scott 1992), and so the basic OAT severely under-samples the experimental region. There exists a variation on the basic OAT, proposed by Morris (1991), which avoids both of these problems, but it will not be discussed here.

The taxonomy of SA methods is complex (Bolado-Lavin and Badea 2008), but one can divide them into two broad categories: local and global. Local methods yield

sensitivity results which are valid only in a small region of the experimental region. The basic OAT approach is a local method because its sensitivity results are reliable only in the vicinity of the fixed values assigned to the inputs. Adjoint methods are also local, because they generally rely on the notion of a derivative or Jacobian of the output with respect to an input (Errico 1997). By contrast, global methods allow an examination of sensitivities across the full experimental region. The main advantage of local methods is their speed and transparency, while the main advantage of global methods is their ability to assess the effect of large changes in the inputs.

Although global methods themselves can be subdivided into finer categories, one class is based on a decomposition of the variance of the output(s) into terms corresponding to the different inputs and their interactions. In this way, such variance-based SA methods can assess the manner in which the uncertainty in an output is apportioned across the inputs, and across interactions between them. Variance-based methods are global in the sense that the sensitivity results do not pertain to any specific value of the inputs; and they are multivariate in that they can assess individual inputs and their interactions. This generality of variance-based SA methods is the reason why it is the main focus of this work.

Variance-based SA involves three ingredients. In the first, classical concepts from analysis of variance (ANOVA) are employed to define measures of sensitivity. The second ingredient involves estimation of conditional expectations in terms of which all of the sensitivity measures are written. The estimation is based on data obtained from running the numerical model for a sample of model parameters; methods of

experimental design are used for deciding what type of sample to take - the third ingredient. All three ingredients are discussed below. The next subsection provides brief reviews of a number of SA applications in the geosciences.[1] It is followed by a section further describing the three ingredients of variance-based SA. Section 3 applies that methodology to the Lorenz '63 model, and compares the results with those obtained from an adjoint-based approach to SA. A summary of the conclusions and their discussion is presented in Section 4. The main goals of the paper are 1) to review and promote the use of variance-based SA methods, and 2) to apply the method to the Lorenz '63 model both as a demonstration of the SA method, as well as gaining a better understanding of the model itself.

1.1 Past Applications of SA

Stein and Alpert (1993) discuss the limitations of OAT analysis in atmospheric numerical models. They show that difference maps do not show interactions between model parameters. For example, they examine the effect of topography and surface fluxes on rainfall, and show that the interaction between the two factors contributes more to rainfall than either of the factors alone.

Alapaty, Raman, and Niyogi (1997) study the effect of five parameters on the boundary-layer structure. Although that work relies on an OAT approach, in a follow-up work Niyogi et al. (1999) extended the analysis by using ANOVA to assess the

[1] For the reader more interested in the technical detail (as opposed to the historical detail), this section may be skipped.

simultaneous effect of the parameters on sensitivity measures.

Murphy et al. (2004) perform an OAT analysis on 29 parameters. The values of the parameters are selected according to expert advice. The forecast quantities considered are the standard deviation (across ensemble) of temperature, precipitation, and pressure. In a similar climate study, Sanderson (2011) performs a comparison of two climate models in terms of the distribution of the values of climate sensitivity (i.e., the change in mean temperature resulting from a doubling of CO_2 concentration), with respect to four parameters. In one model they consider four parameters, each sampled at three levels; in statistics, such a design is called a 3^4 factorial design.

Mishra, Cherkauer, and Bowling (2010) examine the effect of 6 parameters on stream flow. They perform an OAT analysis, followed by a 2^8 factorial analysis.

The idea of sampling different model parameters and/or different initial conditions arises also in ensemble prediction systems, where the main focus is on sampling the initial conditions so as to maximize the variance (across the ensemble) of the forecasts. There exist different methods for generating these samples. Some are designed to optimize the manner in which perturbations grow in time; methods based on singular and bred vectors belong to that class. Magnusson, Leutbecher, and Källén (2008) compare these two methods for sampling initial perturbations. For the purpose of sampling model parameters, however, one often considers random or systematic sampling techniques. Hacker et al. (2011), for example, use a variant of the latter, called latin hypercube sampling, to set the values of four model parameters in a mesoscale

model. They also perform two types of SA; in the first, they examine the effect of each of the four parameters on the mean absolute difference between a forecast where the model parameter is set to a minimum value and another forecast where the model parameter is set to a maximum value. In other words, they follow an OAT, 2^4 factorial design. They also study the effect of the model parameters on the mean and variance of the forecasts across the spatial domain.

Many of the basic ideas in SA arise in the calibration of ensembles, as well. For example, Golaz et al. (2007) consider 10 parameters in a single-column model parameterizing boundary layer clouds. They take simple random samples from a uniform distribution centered on specified initial values of the parameters, and then optimize the parameters by minimizing a squared-norm cost function. They highlight how such an analysis can help in identifying structural errors in the model. Although similar to SA, this type of analysis is different both in the ultimate goal of the analysis and in the technical implementation. As an example of the latter, note that calibration requires a cost function comparing model output with observations, but SA does not require observations at all.

The impact of observations can also be assessed through SA methods (Torn and Hakim 2008). Hakim and Torn (2008) propose a method, called ensemble synoptic analysis, which offers yet another means of performing SA suitable for assessing sensitivity to initial conditions. Gombos and Hansen (2008) apply the ensemble synoptic analysis method to potential vorticity forecasts from Weather Research Forecasting (WRF) Model, and compare the results of that statistical technique with those of a

dynamic approach due to Davis and Emanuel (1991). Ancell and Hakim (2007) compare the ensemble-approach to SA with the more traditional adjoint-based approach, and show that the two methods are equivalent when the initial conditions are spatially uncorrelated. Beyond this simple comparison, which is proven analytically, further comparisons are complex, and so the two methods have their respective advantages and disadvantages.

A variance-based SA of a numerical model in the geosciences is performed by Zhao and Tiede (2011). The numerical model examined there involves five parameters, and the response is changes in gravity measured at the Earth's surface in the vicinity of a volcano; the model parameters are randomly sampled from a uniform distribution.

The "data" generated by all such studies are called computer data, and are further discussed in Section 2.3. One of the earliest general treatments of computer data, at least in meteorological circles, dates back to 1993 (Bowman, Sacks, and Chang 1993). That work demonstrates the importance of proper sampling in assessing the effects of five parameters in a general circulation model.

2 Variance-Based SA

The specific formulation of variance-based SA presented here follows that of Oakley and O'Hagan (2003). The foundations of the variance-based approach are based on two mathematical facts. The first is the variance-decomposition formula, also known

as the law of total variance (Weiss 2005, p. 385):

$$V[y] = V[\,E[y|x_i]\,] + E[\,V[y|x_i]\,] \ , \tag{1}$$

where $E[.]$ and $V[.]$ denote expected value and variance, respectively. $E[y|x_i]$ and $V[y|x_i]$ denote the conditional expected value and conditional variance, respectively, of the output, given the inputs x_i; here, i refers to the i^{th} input. Intuitively, this decomposition states that the total variance in y, $V[y]$, can be written as the sum of two terms, one measuring the variance "between" the conditional means, and the other measuring the mean of the conditional ("within") variances.

The second fact is often known as the High-Dimensional Model Representation (Santner, Williams, and Notz 2003; Sobol' 1993); it states that any function of the type, $y = \eta(x_1, x_2, ..., x_n)$, can be decomposed as follows:

$$y = \eta(x_1, x_2, ..., x_n) = E[y] + \sum_{i}^{n} \ z_i(x_i) + \sum_{i<j} z_{ij}(x_i, x_j) + ... \ , \tag{2}$$

where

$$z_i(x_i) \ = \ E[y|x_i] - E[y] \ , \tag{3}$$

$$z_{ij}(x_i, x_j) \ = \ E[y|x_i, x_j] - E[y|x_i] - E[y|x_j] + E[y] \ . \tag{4}$$

The $z_i(x_i)$ are referred to as main effects, and the $z_{ij}(x_i, x_j)$ are called the first-order interactions. The ... indicates that there exist higher-order interaction terms in the expansion, but here they are assumed to be relatively small. The approximation in (2) is adequate for a wide range of functions, but has been well-studied for piece-wise continuous functions (Chowdhury, Rao, Prasad 2008). Intuitively, η represents the

function mapping the model parameters to the forecast quantity of interest. It is important to point out that the computation of the expected values and variances requires the joint probability distribution of all the inputs. As a result, even a main effect computation for a given input generally involves the other inputs.

2.1 Measures of Sensitivity

The measure of importance for an input is a user-dependent quantity, but a few common measures natural to the formulation of the variance-based methods are as follows (Oakley and O'Hagan 2004). One natural measure gauges the expected reduction in the variance of the output, given an input: $E[\,V[y]-V[y|x_i]\,]$. This measure, denoted V_i, is equal to $V[y]-E[\,V[y|x_i]\,]$, which according to (1) can be written as

$$V_i = V[\,E[y|x_i]\,] \quad . \tag{5}$$

Similarly, the expected reduction in the variance of the output, given two inputs, x_i and x_j, is

$$V_{i+j} = V[E[y|x_i, x_j]] \quad . \tag{6}$$

The quantity V_{i+j} measures the sensitivity of the output with respect to *both* x_i and x_j. From (2)-(6) it follows that $V_{i+j} = V_i + V_j + V[z_{ij}(x_i, x_j)]$, and so, a measure that gauges the level of interaction between x_i and x_j (at least when $z_i(x_i)$ is independent of $z_j(x_j)$) can be defined as

$$V_{ij} = V[z_{ij}(x_i, x_j)] \; . \tag{7}$$

Another useful quantity measures the uncertainty remaining in the output, given

all inputs, except x_i. For example,

$$V_{T1} = V[y] - V[E[y|x_2, x_3, ...]] \ , \tag{8}$$

measures the remaining/unexplained variance in y after all inputs have been fixed, except x_1. Traditionally, the V_i and V_{Ti} measures are converted to proportions, as follows:

$$S_i = V_i/V[y] \quad \text{and} \quad S_{Ti} = V_{Ti}/V[y] \ , \tag{9}$$

where S_i and S_{Ti} are called the main effect index, and the total effect index, respectively. Although they measure different facets of the importance of the i^{th} input, they do not capture interactions between the inputs. Therefore, it is important to supplement S_i and S_{Ti} with V_{i+j} or V_{ij} for a more complete assessment of sensitivity. Table 1 provides a summary of these measures and their meaning.

In special cases, these sensitivity measures can be related to other, common measures of sensitivity. For example, it can be shown that for $y = \eta(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, if x_1 and x_2 are independent and centered (i.e., with $E[x_i] = 0$), then $V_i = \beta_i^2 V[x_i]$, and $S_i = \beta_i^2 V[x_i]/V[y]$. In other words, V_i and S_i are directly related to the "regression" coefficients β_i. Other sensitivity measures have more complex relations to the β_i parameters (Marzban 2011).

2.2 Estimation of $E[\text{output}|\text{input}]$

All of the measures in Table 1 can be computed from the z_i and z_{ij} defined in (3) and (4), which are written in terms of conditional expected values. A great deal

of the work in variance-based SA methods is focused on efficient and accurate ways for estimating these conditional expectations from data. The data themselves are generated by evaluating the function $\eta(.)$ for some set of x_i values. The choice of the x_i values is a complex issue belonging to the realm of experimental design, described in the next section.

The methods for estimating the conditional expectations are varied, but they can be broadly divided into Monte Carlo methods (Sobol' 1993; Cukier et al. 1973; Saltelli et al. 2008, 2010), and methods based on emulation, or meta models (Dusby 2008; Oakley and O'Hagan 2004; Rougier 2008; Rougier et al. 2008). In the former, the conditional expected values are expressed as their defining integral form, which are then evaluated using Monte Carlo techniques. The latter methods aim to develop a statistical model that approximates $y = \eta(.)$. The resulting statistical model is said to emulate the "black box" model $\eta(.)$. The emulator is then employed to estimate the conditional expectations.

The development of an emulator is a complex and sophisticated procedure. Oakley and O'Hagan (2004) develop a Gaussian Process emulator, whose mathematics is similar to Kriging (Chaloner and Verdinelli 1995; Sacks, Schiller, and Welch 1989; Sacks, et al. 1989; Welch et al. 1992). More recently, these methods have been extended to allow multiple outputs (Rougier 2008; Rougier et al. 2008).

The estimation method adopted in this paper is a simple emulation approach based on cubic polynomial regression. Recall that a least-squares fit to data of the

form $(y, x_1, x_2, ...)$ has the property that the fitted value estimates the conditional expected value of the response y, given the predictors $x_1, x_2, ...$ (Bishop 1996; pages 201-202). Specifically, here, $E[y|x_1]$ is estimated by fitting a cubic polynomial through data on x_1 and y. Similarly for $E[y|x_2]$, and $E[y|x_1, x_2]$, etc.

2.3 Experimental Design

Experiments involving numerical models generate data which have no experimental error. In other words, a unique set of values for the inputs will always produce the same output. Such data are called *computer data,* and experiments involving computer data are often called *in silico* - in contrast to *in vitro* or *in vivo* experiments performed, respectively, in a laboratory tube or in a living body. Again, the distinguishing characteristic of computer experiments is the absence of experimental error (Fang, Li, and Sudjianto 2006; Santner, Williams, and Notz 2003). Consequently, the analysis of computer data is somewhat different from that of "real" data. The framework for analyzing computer data is well-established (Santer et al. 2003).

In a computer experiment all of the variability of the response is due to variability in the factors. In other words, for computer data one has $V[y|x_1, x_2, ..., x_n] = 0$, where n is the number of inputs. The variability in y that is utilized to define all the sensitivity measures in Table 1 originates from the variability of the inputs that are *not* fixed. Either way, there *is* variability in y even for computer data, and so care must be taken in sampling, because the precision of the sensitivity measures depends

on the choice of the sampling scheme.

Discretizing the inputs on a grid is a form of sampling, but it is inefficient. For example, consider a 2-dimensional grid discretizing the space spanned by two inputs. Then, any two grid points have at least one input in common. As such, the empirical region is not adequately sampled. An alternative to a grid is random sampling of the experimental region. There exists a wide variety of sampling techniques (Douglas 2005), but only two are considered here: *Simple Random Sampling* (SRS) and *Latin Hypercube Sampling* (LHS). The latter is a systematic sampling method which belongs to a class of sampling techniques called *space-filling*. The pros and cons of all of these sampling methods have been thoroughly examined (Cioppa and Lucas 2007; Fang, Li, and Sudjianto 2006; Santner, Williams, and Notz 2003; Urban and Flicker 2010).

SRS has the desirable property of leading to the most precise estimate of the mean, if the population is homogeneous. For example, in sampling a continuous quantity such as height, if the population of heights has no clusters that distinguish different height characteristics, then a mean height computed from an SRS has the smallest variance across multiple samples (if multiple samples were taken). By contrast, when the population is not homogeneous, LHS is designed to yield no-less precise estimates than SRS.[2] The reason for this difference between the two sampling schemes is that SRS has a tendency to generate clusters. This clustering of data is not of concern when the population is homogeneous, but otherwise results in suboptimal estimates.

[2]It can be shown that LHS cannot lead to less precise estimates than SRS. In fact, if the function $\eta(.)$ is monotonic, then this can be proved analytically (McKay, Beckman, and Conover 1979).

LHS, however, is designed to scatter the sample across the full experimental region. This latest property is the reason why LHS is said to be *space filling* (Cioppa and Lucas 2007; Fang, Li, and Sudjianto 2006; Santner, Williams, and Notz 2003).

Figure 1 shows an example of a SRS (open circles) and a LHS (filled circles) taken from a 2-dimensional experimental region. This specific realization is uncharacteristic in that the SRS circles clearly cluster together, while the LHS circles do not. However, it serves to demonstrate that SRS may lead to clusters, but LHS cannot. Indeed, by design, no two cases in the LHS have the same values of x_1 and x_2.

3 Lorenz '63

Marzban (2011) considers a few examples which can be solved analytically. One example in which the function representing the "black box" model is not available in analytic form is the Lorenz '63 model (Lorenz 1963). It is defined as

$$\begin{aligned} dX/dt &= -s(X-Y), \\ dY/dt &= rX - Y - XZ, \\ dZ/dt &= XY - bZ, \end{aligned} \tag{10}$$

where the model parameters are s (the Prandtl number), r (the Raleigh number), and b, the latter being a function of the wavenumber. The state space variables X, Y, and Z measure the intensity of convective motion, and horizontal and vertical temperature variation, respectively (Bellomo and Preziosi 1995).

In terms of the aforementioned "black box," the inputs are the model parameters s, r, and b. In this paper, the outputs are not the state space variables X, Y, and Z, but rather the mean value of these quantities over a 20-time-step forward integration of the Lorenz equations. These means are denoted X_{mean}, Y_{mean}, and Z_{mean}.

In order to obtain a visual sense of the relationship between the outputs and the inputs, 50 equally-spaced values are selected for each parameter, the Lorenz equations are integrated forward in time steps of 0.02, and $X_{mean}, Y_{mean}, Z_{mean}$ are computed. The empirical region (i.e., range of parameters) includes the default values, and is selected to yield a reasonably rich and complex relationship between the outputs and the inputs. That relationship (for only Z_{mean}) is shown in Figure 2.

Although all three parameters are varied, the various panels in this figure show different cross-sections of the relationship. The top/left panel shows Z_{mean}, in different shades of gray (white = low, black = high), versus s and r, with the b-parameter set to its default value of 8/3. The top/right panel shows a different perspective of the same relationship; the r parameter is shown along the x-axis, and so, the resulting scatter in the figure corresponds to different values of the s parameter. It can be seen that Z_{mean} is mostly monotonically increasing with s and r, with a slight nonlinearity in the extremes where s and r are both either low, or high.

The middle panels show the dependence of Z_{mean} on s and b, with r set to its default value of 28. The behavior here is mostly monotonically decreasing with b; in fact, that dependence visually overwhelms the dependence on s noted in the top

panels. The bottom panels show Z_{mean} versus r and b, with s set to its default value of 10. This time, the relationship is monotonically decreasing with b, but with significant variability due to r.

In short, a wide range of behavior can be seen over the specific range of parameters selected. Beyond this empirical region the relationships are more complex, even consisting of discontinuities. Such complexity is not an obstacle in variance-based SA methods, because the emulators often used are highly nonlinear statistical models capable of modeling a wide range of functional behavior. Here, for simplicity, the model employed to emulate the relationships shown in Figure 2 is a cubic polynomial; there is no evidence from Figure 2 to suggest a more complex model.

The "data" shown in Figure 2 pertain to 50^3 points in parameter space because 50 different values of each parameter are selected. Such a large sample is possible here only because the Lorentz '63 model is relatively simple. A large sample is chosen only to lead to graphs that are visually informative, such as those in Figure 2. The data for performing SA are not obtained by systematically incrementing all the inputs, for that would be computationally prohibitive for most realistic models. Thus, even though the Lorenz model is sufficiently simple to allow a brute-force approach, the SA is performed on random samples (according to SRS and LHS) of size 50 taken from the experimental region spanned by the parameters s, r, and b.

In order to compare the two sampling schemes, 100 samples are taken. The resulting sampling variability of all the sensitivity measures is displayed through boxplots.

It is important to point out that in a realistic application of the variance-based SA, this step of taking multiple samples is unnecessary. The only reason it is done here is to allow a comparison of the two sampling schemes.

There exists another source of variability in the Lorenz model, namely initial conditions. Here, for simplicity, and remaining focused on the illustration of the variance-based SA method, every time a sample is taken from the experimental region, the initial conditions are fixed at their default values, $(X, Y, Z) = (-14, -13, 47)$. A similar analysis but where the initial conditions are also randomly selected has been performed. The results are not shown here, because the only difference is "larger" boxplots resulting from the additional variability due to initial conditions.

Figure 3 shows the sensitivity measures for Z_{mean}; results for X_{mean} and Y_{mean} are not shown here. The boxplots summarize the corresponding distributions resulting from 100 trials (i.e., 100 different samples of size 50 taken from the empirical region). The wide/white boxplots correspond to SRS, and the narrow/gray boxplots are for LHS. Comparing the two boxplots across all the panels in Figure 3, it is clear that LHS yields no-less precise estimates (than SRS) for all of the sensitivity measures; this is evident in the fact that the gray boxplots are generally "shorter" than the white boxplots for many of the sensitivity measures.

The top/left panel shows the distribution of the V_i measures (5) for s, r, and b. Evidently, Z_{mean} is most sensitively dependent on the b parameter, followed closely by the r parameter. The parameter s appears to have no effect at all. However, it would

be a mistake to dismiss s as an important input, because it may be important only in the presence of the other parameters. The measures $V_{s+r}, V_{s+b}, V_{r+b}$ (6) must then be consulted; they are shown in the top/right panel. Comparing this panel with the top/left panel, it is difficult to determine if s is indeed useless, because, for example V_{s+r} (in top/right panel) appears to be comparable with $V_s + V_r$ (in top/left panel). In order to better isolate the effect of s, sensitivity to interaction terms V_{sr}, V_{sb}, V_{rb} (7) may be examined; they are shown in the left/middle panel. There appears to be evidence for a weak interaction between r and b, but it is not clear if s interacts with r or with b.[3] The remaining panels show the "total" sensitivity measures (8, 9) and confirm that that Z_{mean} is most sensitively dependent on b and r. These conclusions are the same regardless of the sampling scheme.

3.1 The Adjoint SA

It is useful to compare the above findings to the results of a more traditional method such as the adjoint method. Hall (1986) has developed a framework wherein the sensitivity of the mean of a state variable, with respect to both initial conditions and model parameters, can be obtained using the adjoint method. Application of that approach to the Lorenz '63 model (with fixed initial conditions) yields

$$\delta Z_{mean} = \int_0^T [\, v_x(Y - X)\, \delta s + v_y X\, \delta r + v_z Z\, \delta b\,]\, dt \tag{11}$$

[3]To settle that question, one would require knowledge of the distribution of the sensitivity measures (i.e., the boxplots) under the null hypothesis of no interaction. See the discussion, below.

where X, Y, Z satisfy (10), $\delta s, \delta r, \delta b$ are perturbations to the model parameters, and δZ_{mean} is the corresponding change in the mean (over time T) of Z.[4] The quantities v_x, v_y, v_z are components of a column vector $\vec{v}$ which must be a solution to

$$N^* \vec{v} = \frac{1}{T} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \vec{v}(t = T) = \vec{0}\,. \tag{12}$$

where N^* is the adjoint of the tangent linear model corresponding to Lorenz '63. The details leading to these results are not shown here, but similar expressions can be derived for δX_{mean} and δY_{mean}. The coefficients of $\delta s, \delta r, \delta b$ in (11) can be thought of as measures of sensitivity of Z_{mean} with respect to parameter perturbations. Note that these coefficients are functions of time T.

Figure 4 shows the logarithm of these sensitivity measures as a function of T, although only $T = 20$ is relevant for comparison with the variance-based sensitivity results. The three line thicknesses - from thin to thick - correspond to sensitivity with respect to $\delta s, \delta r$, and δb, respectively. It can be seen that the curve corresponding to δb is consistently above the δr curve, which in turn is higher than the δs curve. This implies that δZ_{mean} is most sensitive to δb, followed by δr, and then δs. And, of course, these results are consistent with the results of the variance-based method.

A more qualitative comparison of the two methods is provided in the next section.

[4] Technically, Hall's method gives the mean of δZ; but under general conditions that quantity is equal to the δZ_{mean}.

4 Conclusions and Discussion

A variance-based method for sensitivity analysis is reviewed and demonstrated on the Lorenz '63 model. The method relies heavily on ideas from analysis of variance, regression modeling, and experimental design (Fang, Li, and Sudjianto 2006; Santner, Williams, and Notz 2003). It is found that the mean of the Z state variable is most sensitively dependent on the b parameter. The sensitivity on the r parameter is comparable, but the s parameter appears to have no effect on Z_{mean} at all. There is also indication of a weak interaction between the r and b parameters. These findings are valid for simple random and latin hypercube samples. It is also seen that the latter scheme leads to more precise estimates for some sensitivity measures, but not for all. The results are generally consistent with an adjoint sensitivity analysis, in so far as the two methods can be compared.

A qualitative comparison of the variance-based and the adjoint method is as follows: The adjoint method has the desirable feature that an analytic expression can be written for the sensitivity (11). However, it has the undesirable property that the sensitivity analysis refers explicitly to a given reference trajectory, i.e., the X, Y, Z appearing in (11). As a result, $\delta s, \delta r, \delta b$ are all perturbations about a reference parameter set; the variance-based method does not have this limitation. Also, the variance-based method naturally allows for an estimate of the uncertainty in the sensitivity measures, e.g., through the boxplots shown in Figure 3; the adjoint-based result (11) does not. On the other hand, in spite of the flexibility of most emulators (in

essentially estimating a response surface) the estimation may simply be wrong. This introduces another source of variability (sometimes called computational uncertainty) in the variance-based method, which does not plague the adjoint approach. The adjoint method of Hall (1986) is naturally suited to estimate δZ_{mean}, but it is not clear how to extend the method to allow the estimation of other quantities of interest, e.g., the change in the maximum of Z, i.e., δZ_{max}. Finally, unlike the variance-based method, the adjoint method in its current form does not allow the estimation of interaction terms; however, that generalization is possible and is currently under investigation. Also underway, are a more thorough assessment of the sensitivity with respect to initial conditions, and an application of variance-based SA to a mature numerical weather prediction model (i.e., COAMPS®)[5].

Here it has been sufficient to emulate the conditional expectations with a cubic polynomial regression. The main reason (for the sufficiency) is that the number of cases generated for "training" the regression model is practically unbounded, and so there is little or no chance of overfitting. In more realistic examples, where 1) the nonlinearity of the $\eta(.)$ model may require more nonlinear functions, and 2) the $\eta(.)$ model is expensive to run, alternative emulators should be used. Storlie et al (2009) compare a wide range of emulation methods; one class of emulators which has gained recent popularity is called gaussian processes regression (Chaloner and Verdinelli 1995; Hsieh 2009; Rougier 2008; Rougier et al. 2008; Sacks, Schiller, and Welch 1989; Sacks, et al. 1989; Welch et al. 1992). These are sophisticated models,

[5]COAMPS is a registered trademark of the Naval Research Laboratory.

whose training requires more effort than polynomial regression, which is the reason why they are not used here.

The analysis performed here has involved only one of the outputs (Z_{mean}), and the approach can be applied to each output, separately; however, others have proposed multivariate methods where any covariance structure existing across the outputs may also be taken into account (Fasso 2006; Oakley and O'Hagan 2004). This approach will be examined in the context of ongoing work with COAMPS®,

Here, it has been sufficient to compare the boxplots of the sensitivity measures, representing sampling variability, only visually. However, a more rigorous treatment will be required if one desires to objectively assess the statistical significance of the results. To that end, it will be necessary to compute/construct the sampling distribution of the sensitivity measures under the null hypothesis of no effect and no interaction. That work is also currently under investigation.

Acknowledgments

Scott Sandgathe and Greg Hakim are acknowledged for valuable discussions. Partial support for this project was provided by the Office of Naval Research grant numbers N00014-01-G-0460/0049 and N00014-05-1-0843.

References

Ancell, B., and G. Hakim, 2007: Comparing Adjoint- and Ensemble-Sensitivity Analysis with Applications to Observation Targeting. *Mon. Wea. Rev.*, **135**, 4117-4134.

Alapaty, K., S., Raman, and D. Niyogi, 1997: Uncertainty in the specification of surface characteristics: A study of prediction errors in the boundary layer. *Boundary-Layer Meteorol.*, **82**, 473500.

Bellomo, N., and L. Preziosi, 1995: *Modelling Mathematical Methods and Scientific Computation*. CRC press. 497 pp.

Bishop, C. M., 1996: *Neural networks for pattern recognition.* Clarendon Press, Oxford, pp. 482.

Bowman, K. P., J. Sacks, and Y-F. Chang, 1993: Design and Analysis of Numerical Experiments. *J. Atmos. Sci.*, **50(9)**, 1267-1278.

Bolado-Lavin, R., and A. C. Badea, 2008: Review of sensitivity analysis methods and experience for geological disposal of radioactive waste and spent nuclear fuel. JRC Scientific and Technical Report. Available online.

Cacuci, D. G., 2003: *Sensitivity and Uncertainty Analysis*. Chapman and Hall/CRC. 285 pp.

Chaloner, K., and I. Verdinelli, 1995: Bayesian experimental design: A review. *Statistical Science*, **10(3)**, 273304.

Chowdhury, R., B. N. Rao, A. M. Prasad, 2008: High dimensional model representation for piece-wise continuous function approximation. *Commun. Numer. Meth. Engn.*, **24** , 1587-1609.

Cioppa, T. and T. Lucas, 2007: Efficient nearly orthogonal and space-filling latin hypercubes. *Technometrics*, **49(1)**, 4555.

Davis, C.A., and K. Emanuel, 1991: Potential vorticity diagnostics of cyclogenesis. *Mon. Wea. Rev.*, **119**, 19291953.

Douglas, C. M., 2005: *Design and Analysis of Experiments*, John Wiley & Sons, 643 pp.

Dusby, D., 2009: Hierarchical adaptive experimental design for Gaussian process emulators. *Reliability Engineering and System Safety*, **94(7)**,1183-1193 .

Errico, R. M., 1997: What is an Adjoint Model? *Bull. Amer. Meteor. Soc.*, **78**, 2577-2591.

Fang K.-T., R. Li, and A. Sudjianto, 2006: *Design and Modeling for Computer Experiments*, Chapman & Hall/CRC, 290 pp.

Fasso, A. 2006: Sensitivity Analysis for Environmental Models and Monitoring Networks. In: Voinov, A., Jakeman, A.J., Rizzoli, A.E. (eds). Proceedings of the iEMSs Third Biennial Meeting: Summit on Environmental Modelling and Software. International Environmental Modelling and Software Society, Burlington, USA, July 2006.

506 Internet: http://www.iemss.org/iemss2006/sessions/all.html
507 ISBN 1-4243-0852-6 978-1-4243-0852-1

508 Golaz, J.-C, V. E. Larson, J. A. Hansen, D. P. Schanen, and B. M. Griffin, 2007: Elu-
509 cidating model inadequacies in a cloud parameterization by use of an ensemble-
510 based calibration framework. *Mon. Wea. Rev.*, **135**, 4077-4096.

511 Gombos, D., and J. A. Hansen 2008: Potential Vorticity Regression and Its Rela-
512 tionship to Dynamical Piecewise Inversion. *Mon. Wea. Rev.*, **136**, 2668-2682.

513 Hall, M. C. G., 1986: Application of adjoint sensitivity theory to an atmospheric
514 general circulation model. *J. Atmos. Sci.*, **43**, 2644-2651.

515 Hacker, J. P., C. Snyder, S.-Y. Ha, M. Pocernich, 2011: Linear and non-linear
516 response to parameter variations in a mesoscale model. *Tellus A*, 63:429444.
517 doi:10.1111/j.1600-0870.2010.00505.x

518 Hakim, G.J. and R. D. Torn, 2008: Ensemble synoptic analysis. *Meteor. Monogr.*,
519 **33**, 147-162.

520 Hsieh, W. 2009: *Machine Learning Methods in the Environmental Sciences: Neural*
521 *Network and Kernels*, Cambridge University Press. 349 pp.

522 Lorenz, E. N., 1963: Deterministic non-periodic flow. J. Atmos. Sci., 20, 130-141.

523 Jimenez, L., and D. Landgrebe, 1998: Supervised Classification in High Dimensional
524 Space: Geometrical, Statistical, and Asymptotical Properties of Multivariate

Data. *IEEE Transactions on System, Man, and Cybernetics*, **28**, Part C, 39-54. 10.1109/5326.661089

Magnusson, L., M. Leutbecher, E. Källén, 2008: Comparison between singular vectors and breeding vectors as initial perturbations for the ECMWF ensemble prediction system. *Mon. Wea. Rev.*, **136**, 4092-4104.

Marzban, C. 2011: Sensitivity Analysis in linear and nonlinear models: A review. Paper presented at the 9th Conference on Artificial Intelligence, at the 91st American Meteorological Society Annual Meeting, Seattle, Jan. 23-27.

McKay, M. D., R. J. Beckman, W. J., Conover, 1979: A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, **21(2)**, 239-245 .

Miller, R. N., 2007: *Numerical Modeling of Ocean Circulation.* Cambridge University Press. 242 pp.

Mishra, V., K. A. Cherkauer, and L. C. Bowling, 2010: Parameterization of Lakes and Wetlands for Energy and Water Balance Studies in the Great Lakes Region. *J. Hydrometeorology*, **11**, 1057-1082.

Morris, M.D., 1991: Factorial sampling plans for preliminary computational experiments. *Technometrics*, **33**, 161174.

Murphy, J.M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations, *Nature*, **430**, 768-772.

546 Niyogi, D. S., S. Raman, and K. Alapaty, 1999: Uncertainty in the specification
547 of surface characteristics, part II: Hierarchy of interaction-explicit statistical
548 analysis. *Boundary-Layer Meteorol.*, **91**, 341-366.

549 Oakley, J. E., and A. O'Hagan, 2004: Probabilistic sensitivity analysis of complex
550 models: a Bayesian approach. *J. R. Statist. Soc.*, **B 66(3)**, 751-769.

551 Richardson, L. F. 2007: *Weather Prediction by Numerical Process (2nd Edition).*
552 Cambridge University Press. 250 pp.

553 Rougier, J. 2008: Efficient Emulators for Multivariate Deterministic Functions.
554 *Journal of Computational and Graphical Statistics,* **17(4)**, 827-843.

555 Rougier, J.C., S. Guillas, A. Maute, and A. Richmond, 2008: Expert knowledge and
556 multivariate emulation: The Thermosphere-Ionosphere Electrodynamics Gen-
557 eral Circulation Model (TIE-GCM). Technical Report 08:22, Statistics Group,
558 University of Bristol.

559 Rushton, K.R., 2003: *Groundwater Hydrology: Conceptual and Computational Mod-*
560 *els.* John Wiley and Sons Ltd. 416 pp. ISBN 0-470-85004-3

561 Sacks, J., S. B. Schiller, and W. J. Welch, 1989: Designs for Computer Experiments.
562 *Technometrics*, **31(1)**, 41-47.

563 Sacks, J., W. J., Welch, T. J. Mitchell, H. P. Wynn, 1989: Design and Analysis of
564 Computer Experiments. *Statistical Science*, **4**, 409-423.

565 Saltelli, A., P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, 2010:

Variance based sensitivity analysis of model output: Design and estimator for the total sensitivity index. *Computer Physics Communications*, **181**, 259270.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, M. Saisana, S. Tarantola, 2008: *Global Sensitivity Analysis: The Primer*, Wiley Publishing, 304 pp.

Sanderson, B.M., 2011: A multimodel study of parametric uncertainty in predictions of climate response to rising greenhouse gas concentrations. *J. Climate*, **24**, 1362-1377.

Santner, T.J., B. J. Williams, and W. I. Notz, 2003: *The Design and Analysis of Computer Experiments.* Springer, 299pp.

Scott, D. W., 1992: *Multivariate Density Estimation; Theory, Practice, and Visualization*. New York: John Wiley & Sons, 325 pp.

Sobol', I. M. 1993: Sensitivity estimates for nonlinear mathematical models, *Mathematical Modeling and Computational Experiments*, v. 1, 407-414.

Stein, U. and P. Alpert 1993: Factor seperation in numerical simulations, *J. Atmos. Sci.*, **50**, 2107-2115.

Storlie, C.B, L.P. Swiler, J. C. Helton, C. J. Sallaberry, 2009: Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. *Reliability Engineering and System Safety*, **94**, 17351763.

585 Torn, R.D., and G. Hakim, 2008: Ensemble-based sensitivity analysis. *Mon. Wea.*
586 *Rev.*, **136**, 663-677.

587 Weiss, N.A., 2005: *A Course in Probability*. AddisonWesley. 789 pp.

588 Welch, W. J., R. J. Buck, J. Sacks, H. P. Wynn, T. J. Mitchell, and M. D. Mor-
589 ris, 1992: Screening, Predicting, and Computer Experiments. *Technometrics*,
590 **34(1)**, 15-25.

591 Zhao, J. and C. Tiede, 2011: Using a variance-based sensitivity analysis for analyz-
592 ing the relation between measurements and unknown parameters of a physical
593 model. *Nonlin. Processes Geophys.*, **18**, 269276.

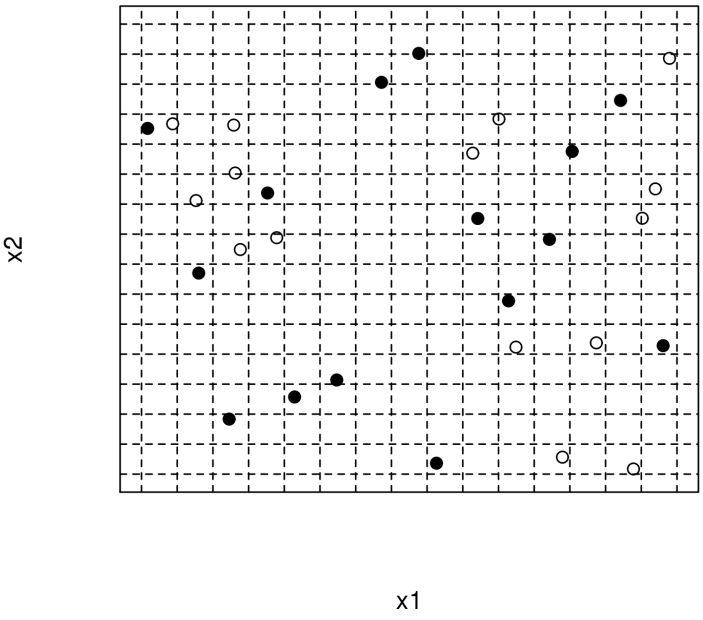
Figure Captions

Figure 1. A SRS (open circles) and a LHS (filled circles). Note that the former is clustered, while the latter "fills" the space.

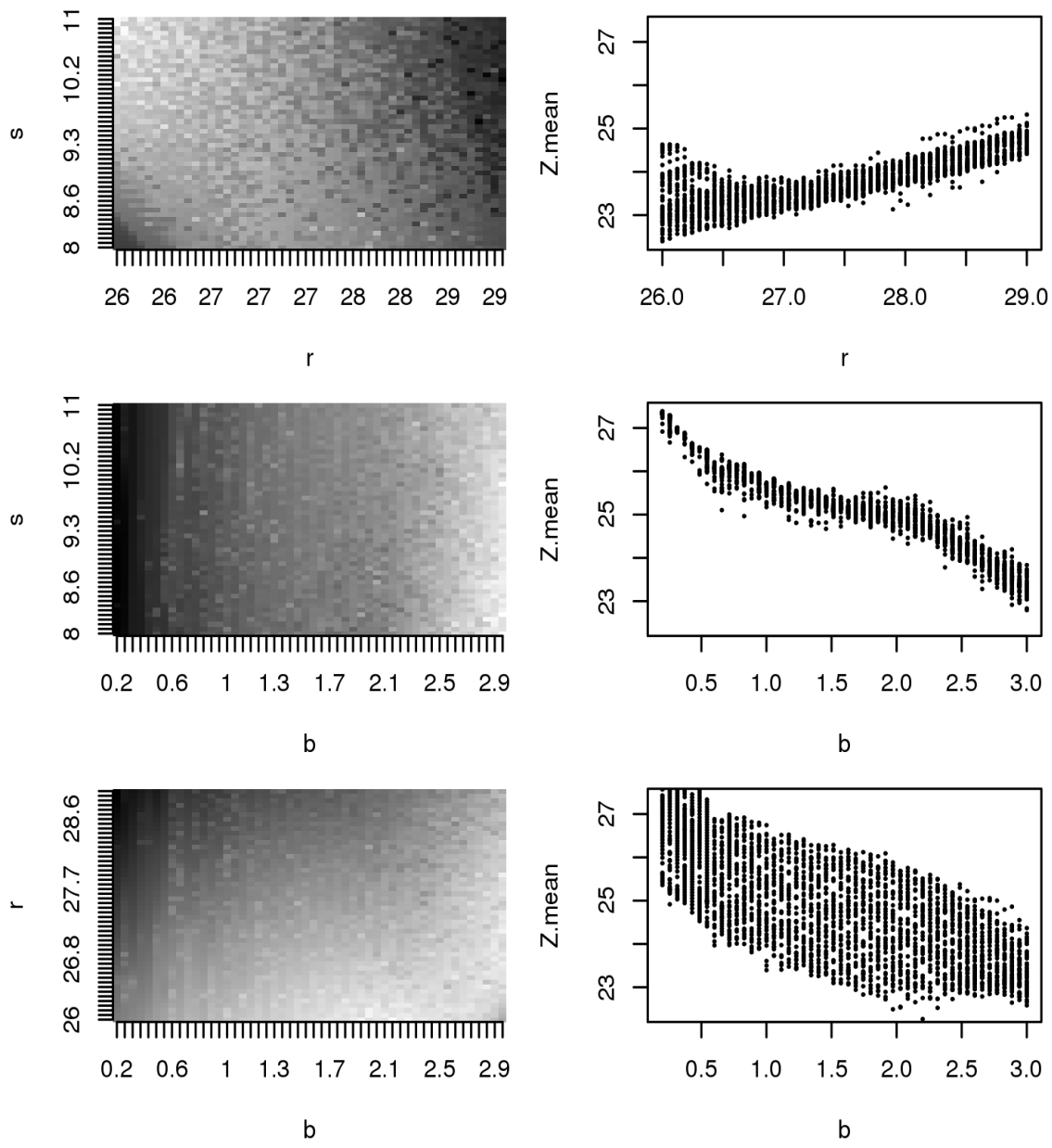
Figure 2. The dependence of Z_{mean} in the Lorenz '63 model on the parameters s, r, and b. The panels in the left column show Z_{mean}, in different shades of gray (white = low, black = high), versus s and r (top), s and b (middle), and r and b (bottom). In each panel, the parameter not shown is set to its default value. The right panels show the same relationships but on 2d scatterplots. See text for explanation.

Figure 3. Variance-based sensitivity measures for the Lorenz '63 model. The boxplots show variability due to sampling variability in the experimental region. The wide/white (narrow/gray) boxplots correspond to SRS (LHS).

Figure 4. Sensitivity measures for the Lorenz '63 model, according to the adjoint method. The thinnest curve gives the sensitivity with respect to s. while the increasingly thicker curves are associated with r and b parameters, respectively.



607 Figure 1. A SRS (open circles) and a LHS (filled circles). Note that the former is
608 clustered, while the latter "fills" the space.



609 Figure 2. The dependence of Z_{mean} in the Lorenz '63 model on the parameters $s, r,$
610 and b. The panels in the left column show Z_{mean}, in different shades of gray (white
611 = low, black = high), versus s and r (top), s and b (middle), and r and b (bottom).
612 In each panel, the parameter not shown is set to its default value. The right panels
613 show the same relationships but on 2d scatterplots. See text for explanation.

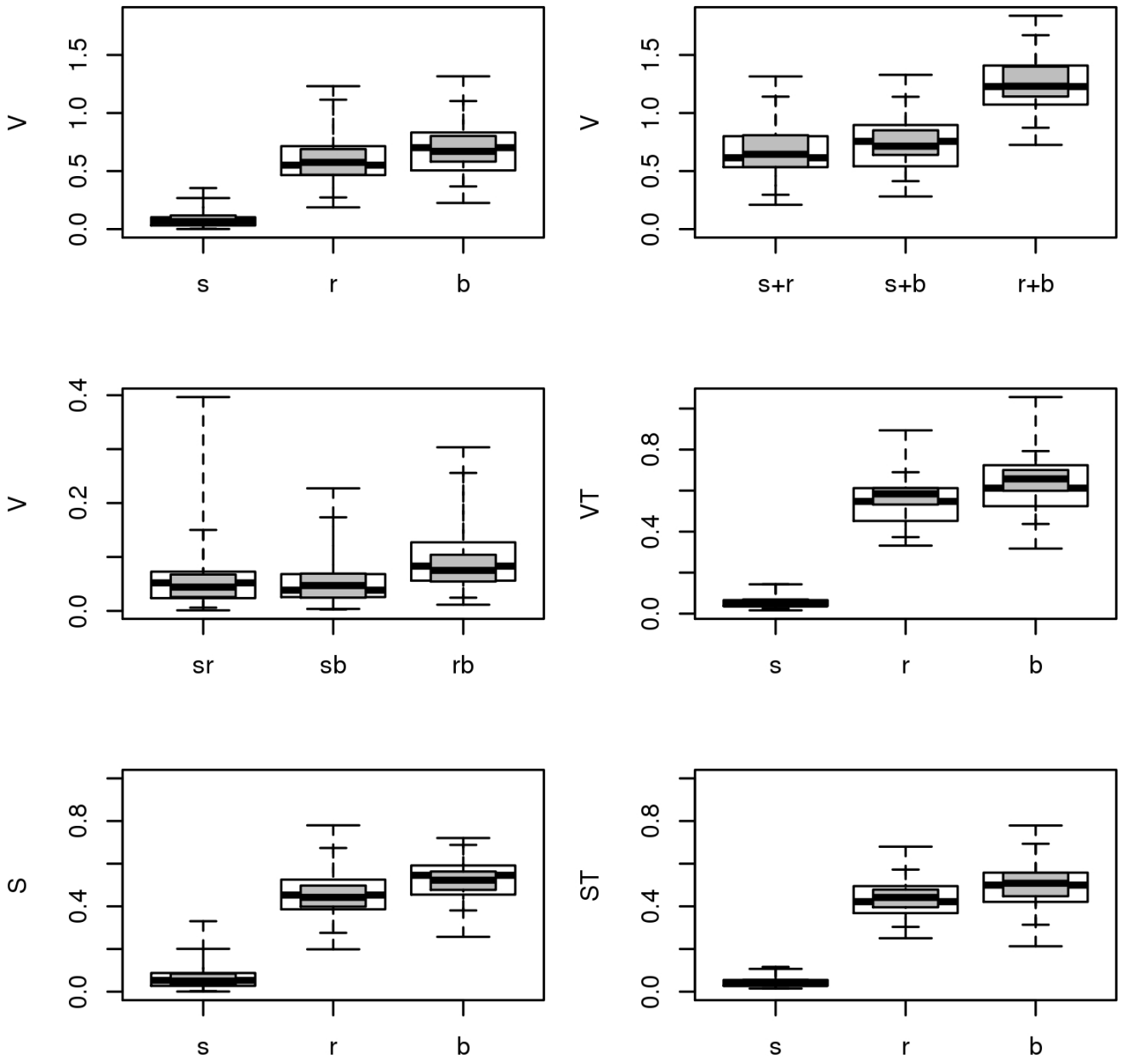


Figure 3. Variance-based sensitivity measures for the Lorenz '63 model. The boxplots show variability due to sampling variability in the experimental region. The wide/white (narrow/gray) boxplots correspond to SRS (LHS).

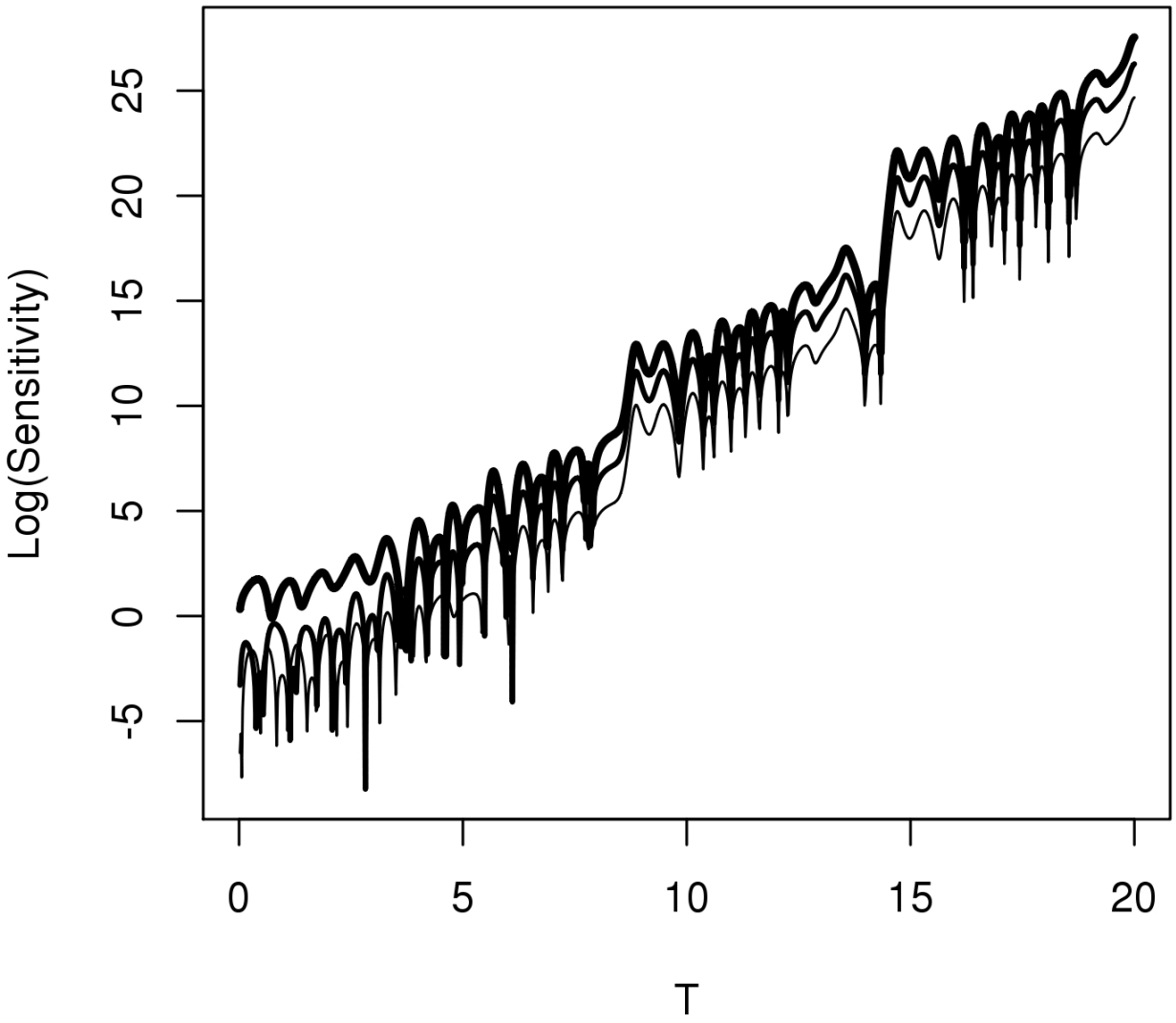


Figure 4. Sensitivity measures for the Lorenz '63 model, according to the adjoint method. The thinnest curve gives the sensitivity with respect to s. while the increasingly thicker curves are associated with r and b parameters, respectively.

Sensitivity measure	Meaning
$V_i = V[\, E[y\|x_i]\,]$	Reduction in uncertainty of y, given x_i
$V_{i+j} = V[E[y\|x_i, x_j]]$	Reduction in uncertainty of y, given x_i and x_j
$V_{ij} = V[z_{ij}(x_i, x_j)]$	Component of V_{i+j} due to interaction between x_i and x_j
$V_{T1} = V[y] - V[E[y\|x_2, x_3, ...]]$	Uncertainty in y remaining, given everything except x_1
$S_i = V_i/V[y]$	Main effect index of x_i
$S_{Ti} = V_{Ti}/V[y]$	Total effect index of x_i

620 Table 1. The sensitivity measures examined here.