5-2014

Sensitivity Analysis of a Leaf Photosynthesis-Stomatal Resistance Model

Gabriel A. Garcia
University of Tennessee - Knoxville, ggarcia4@utk.edu

Sensitivity Analysis of a Leaf Photosynthesis-Stomatal Resistance Model

Gabriel Alberto Garcia
University of Tennessee, Knoxville

A thesis submitted in fulfillment of the Honors Thesis requirement for the Degree of Bachelor of Science in Mathematics with Honors Concentration

Advisor: Dr. Xiaopeng Zhao
Department of Mechanical, Aerospace, and Biomedical Engineering
University of Tennessee, Knoxville

April 2014

Acknowledgement

I would like to thank my advisor Dr. Xiaopeng Zhao who has been supporting me since 2009 when I first became involved in research during high school. I am grateful for the amounts of knowledge about research I have gained from working alongside Dr. Zhao for the past five years in projects from areas such as electrocardiography to climate models. I also appreciate all the advice and help from Dr. Zhao which was greatly used in completing this work.

I would also like to thank Dr. Dali Wang who is a research scientist of the Environmental Sciences Division & Climate Change Sciences Institute at Oak Ridge National Laboratory and Mr. Yang Xu who is currently a Ph.D. student in the Department of Geography at UTK. Dr. Wang and Mr. Xu provided the computational model used in this work, answered many of my questions, and provided technical assistance. Dr. Wang first gave the problem statement of this thesis to Dr. Zhao who then gave me the privilege to undertake the project.

Lastly, I especially thank my father, mother, and brother for all their support throughout my whole life including my time at the University of Tennessee. I would not have been able to make it this far without them. They are true inspirations to me.

Table of Contents

Abstract

One of the biggest problems that we face in the 21st century is climate change, especially global warming. The Community Land System Model (CLM) helps scientists understand how human and vegetation can affect the climate. An important task at the moment is to link measurements collected at a site with results computed by the CLM components. This project will investigate the influences of critical parameters to photosynthesis by carrying out sensitivity and uncertainty analyses of a leaf photosynthesis-stomatal resistance model that is utilized by the CLM. Such techniques will allow us to understand how the variation in the output parameters can be related to changes in input parameters. SimLab software was used for Monte Carlo analysis and generation of sensitivity indices through the methods of Extended FAST, Sobol, regression, and Morris. A ranking of influential parameters can then be determined based on each method. The results shed light on the influential significance of vegetation temperature and other parameters to photosynthesis.

Chapter I

General Introduction

Understanding the behavior of many natural phenomena such as storms, diseases, earthquakes, and chemical reactions has been marked by the use of mathematical models in recent times. Besides their usefulness in analyzing phenomena, mathematical models are also important because they can make valuable predictions. Since mathematical models are built under assumptions that are more idealized than in nature, mathematical models must be examined carefully. In 1987, the statistician George Box infamously warned that "all models are wrong, but some are useful" [1].

In this work, one aspect of a mathematical model for climate change is examined through parametric sensitivity analysis. Climate change is a very important topic of investigation in the 21st century as global warming is a concern among scientists. Scientists from across various disciplines such as meteorology, chemistry, biology, computer science, and engineering have all collaborated on understanding global warming and possible ways to prevent it from intensifying. One such outcome of these collaborations has been the Community Land Model (CLM) which has various components. This work will focus on a leaf photosynthesis and stomatal resistance component of CLM.

As mentioned before, sensitivity analysis is the method of study for this work. Sensitivity analysis is about analyzing a model to improve it. More specifically, sensitivity analysis involves changing the parameter values in a model over a series of model trials and measuring how much the output values changes due to the changes in the parameter values. Sensitivity analysis can then be utilized to see which are the most important parameters (and equally the least important ones) in the model. The model can then be simplified by removing non-influential parameters for example. Besides simplification, sensitivity analysis can assist in model validation, distribution of resources, promotion of new means of collecting data or performing the experiment, exemplification of unrealistic model behavior, and identification of useful model assumptions [2]. Sensitivity analysis can be done in several ways due to

different theoretical approaches in calculating sensitivity measures of the parameters. Many of these approaches require frameworks involving probability and statistics which will also be explored. The final task is then to perform sensitivity analysis on the leaf photosynthesis-stomatal resistance component of CLM and identify which are the more sensitivity (important) parameters.

Chapter 2

Community Land Model

The National Center for Atmospheric Research (NCAR) has developed many computational models for the global atmosphere over the past three decades starting with the Community Climate Model (CCM) in 1983. In 1994, they improved CCM through the Climate System Model (CSM) which incorporated more model components for the atmosphere, land surface, ocean, and sea ice. Afterwards, the Department of Energy (DOE), the National Aeronautics and Space Administration (NASA), and the National Science Foundation (NSF) began to provide more funding for CCM with the goal of encouraging more participation of the scientific community. In 1996, CCM was renamed as the Community Climate System Model (CCSM). One particular experiment from CCSM in 1998 predicted that the concentration of atmospheric carbon dioxide would increase three-fold in 125 years. Today, the model is still being used and improved.

The climate is a sophisticated system that is influenced by physical, chemical, and biological processes that occur throughout the atmosphere, water, and land. Furthermore, these processes can either be influenced by natural causes or human causes. Increasing scientific understanding and computational capabilities are thus both crucial in developing a more accurate model of the climate. Therefore, NCAR has created CESM as an "evolving model" for the greater scientific community. With a more refined model, national and international policies regarding climate change can be better made [3].

Biogeochemical processes occurring on land is one of the most significant factors in climate change. The Community Land Model (CLM) was created by NCAR to be the land (terrestrial) component model for CESM. CLM is associated with the quantification of ecological climatology which involves studying how nature and human beings affect vegetation which in turn affects climate. There are several aspects considered within CLM such as surface heterogeneity, biogeophysics, hydrologic cycle, biogeochemical cycles, ecosystem dynamics, and human impact [4].

Surface heterogeneity is a factor that takes into consideration that land is not uniform. A piece of land can be modeled as a grid which then can be divided into sub-grids known as landunits as shown in Figure 1. The main landunits are glacier, wetland, vegetated, lake, and urban. The vegetated landunit is further sub-divided into patches of plant functional types such as tropical broadleaf deciduous tree and C_4 grass. Each plant function type patch has an associated leaf and stem area index and canopy height [5]. Quantities relating to energy and water balance are calculated for each patch at every time step so at the end of the calculations each patch has prognostic variables. The patches do not interact with each other in a direct manner [6].

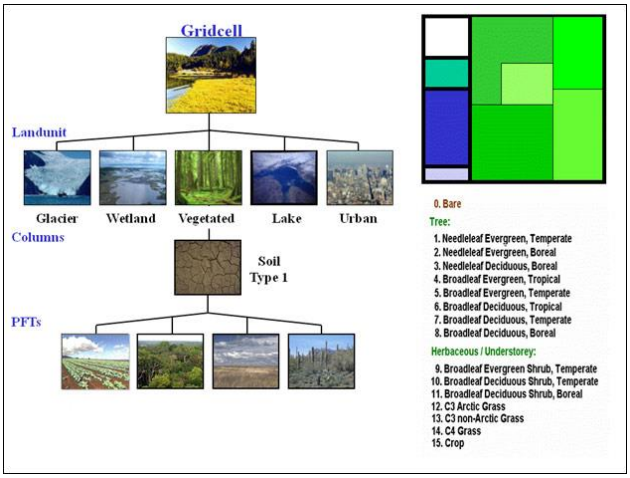
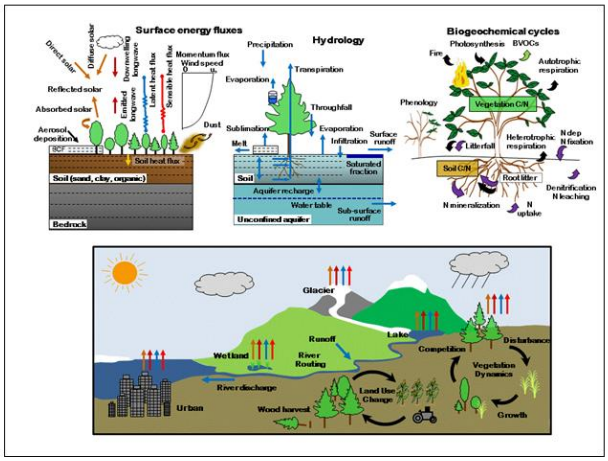


Figure 1. There exist 5 types of main sub-grids (named landunits) that partition a land surface grid. The vegetated landunit can be re-divided into patches of plant functional types [5].

CLM is comprehensive in describing many processes such as absorption, reflection, and transmittance of solar radiation, lake temperatures, stomatal physiology and photosynthesis, and carbon-nitrogen cycling. Figure 2 contains schematic diagrams of the various aspects of CLM along with a list of land surface processes represented in the CLM [4, 7].



- Vegetation composition, structure, and phenology
- Absorption, reflection, and transmittance of solar radiation
- Absorption and emission of longwave radiation
- Momentum, sensible heat (ground and canopy), and latent heat (ground evaporation, canopy evaporation, transpiration) fluxes
- Heat transfer in soil and snow including phase change
- Canopy hydrology (interception, throughfall, and drip)
- Snow hydrology (snow accumulation and melt, compaction, water transfer between snow layers)
- Soil hydrology (surface runoff, infiltration, redistribution of water within the column, sub-surface drainage, groundwater)
- Stomatal physiology and photosynthesis
- Lake temperatures and fluxes
- Dust deposition and fluxes
- Routing of runoff from rivers to ocean
- Volatile organic compounds emissions
- Urban energy balance and climate
- Carbon-nitrogen cycling
- Dynamic landcover change
- Dynamic global vegetation

Figure 2. The CLM model incorporates many components that take account of surface energy fluxes, hydrology, biogeochemical cycles, vegetation dynamics, and anthropogenic impact on the land. Furthermore, specific processes of these components are listed. Note that SCF is snow cover fraction, BVOC is biogenic volatile organic compounds, C/N is carbon and nitrogen. Lastly, the black arrow in the Biogeochemical Cycles diagram refers to carbon flux while the purple arrow refers to nitrogen flux [7].

CLM has been tested rigorously against other developed land models and thus has gone through four stages of improvement. As of spring 2014, version 4.5 is the current version of CLM [4]. When the testing is performed, the model results are compared with observational data that has been collect over various years and from different regions of the world. It has been shown that the CLM performs better in regard to simulation of many processes than the other developed land models [6]. CLM uses a vast array of model parameters, many of which involve uncertainties. CLM consists also of various methods known as subroutines with associated parameters that calculate certain quantities. The subroutine of CLM and its associated parameters of most importance to this work are those of leaf photosynthesis and stomatal resistance which will be the topic of the next chapter.

Chapter 3

Leaf Photosynthesis and Stomatal Resistance

Photosynthesis is the process by which plants convert light energy into chemical energy that is stored in carbohydrate molecules (primarily as glucose, $C_6H_{12}O_6$) as shown in Figure 3. The carbohydrate molecules can later be broken down to release energy as fuel for the plant's metabolism. The carbohydrate molecules are synthesized from carbon dioxide (CO_2) and water (H_2O). Oxygen gas (O_2) is then released as a waste product. During photosynthesis, light energy usually from the sun is captured by organelles known as chloroplasts found in leaf cells. Chloroplast contains green chlorophyll pigments and thus is the reason why vegetation is mainly green. The biochemistry of photosynthesis is actually more detailed and involves several of stages and pathways and thus will not be discussed in this work. However, photosynthesis can be summarized by the chemical equation:

$$6\ CO_2 + 12\ H_2O + \text{Light} \rightarrow C_6H_{12}O_6 + 6\ O_2 + 6\ H_2O.$$

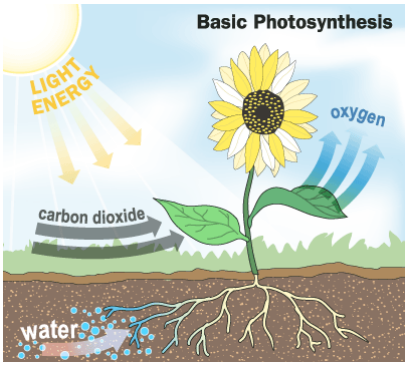


Figure 3. A simplified representation of how photosynthesis works. Note that the carbohydrates are not shown as they are produced within the plant [8].

There exist pores on the leaves by the name of *stomata* by which carbon dioxide flows in while the water and oxygen flow out. Because plants require water to live, plants can retain water when sufficient carbon dioxide needs are met through closing off its stomata. Figure 4 depicts actual stomata on a leaf and a demonstration of how they work.

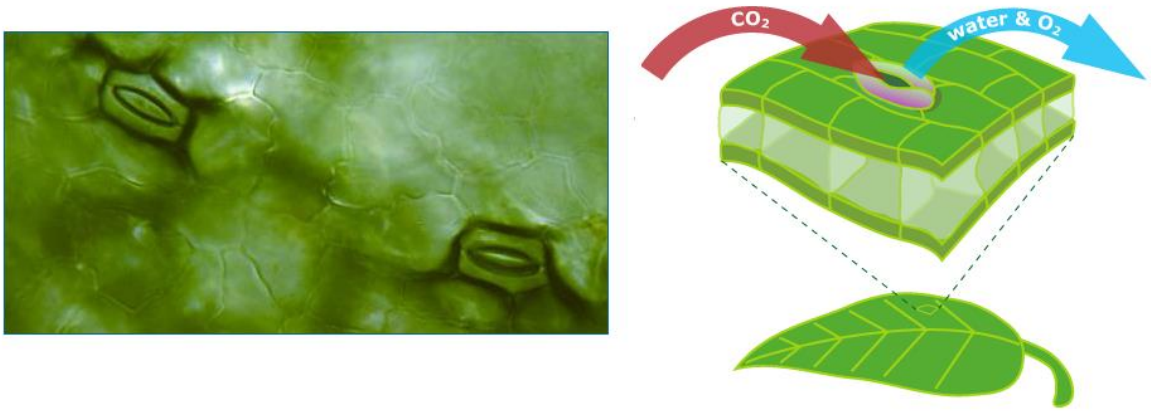


Figure 4. The left picture is a photograph of two actual stomata while the right picture demonstrates that carbon dioxide enters the stomata while water and oxygen gas leave the stomata [9].

In the CLM, stomatal resistance and leaf photosynthesis are both coupled together in a model that is inspired by the C_3 photosynthesis model of Farquhar et al. (1980), the C_4 model of Collatz et al. (1992), and the leaf stomatal conductance model of Ball (1998) [10, 11, 12]. The rate of passage of carbon dioxide entering the stomata is defined as the stomatal conductance. Stomatal conductance can also be defined as the rate of passage of water vapor exiting the stomata. The inverse of stomatal conductance is stomatal resistance which is directly related to the boundary layer resistance of the leaf and the absolute gradient of water vapor from the leaf to the atmosphere. The rate for leaf photosynthesis is known as the gross photosynthetic rate (denoted by A) is the rate at which carbon dioxide is assimilated in the leaf and converted into carbohydrates. The gross photosynthetic rate can be quantified as the minimum of three limiting rates. The first rate is associated with the efficiency of the RuBisCO enzyme involved in photosynthesis (Rubisco limited rate of assimilation, denoted by w_c). The second rate is associated with the amount of photosynthetically active radiation (PAR) that is absorbed by chlorophyll (light-limited rate of assimilation, denoted by w_e). Finally the third rate is associated with the products of photosynthesis and how they are used (Carbon compound export limitation, denoted by w_s). Therefore, the equation for the gross photosynthetic rate is $A = \min(w_c, w_e, w_s)$. Sellers et al. (1995) includes functional definitions of the three rates [13]. Another

quantity can be defined as the net gross photosynthetic rate (A_n) which is $A_n = A - R_d$ where Rd is the leaf respiration rate. The leaf stomatal conductance model of Ball (1998) is combined with the above photosynthetic model to give the equation $\frac{1}{r_s} = m\frac{A_n}{C_s}\frac{e_s}{e_i}P_{atm} + b$ where r_s is leaf stomatal resistance, m is the plant functional type dependent parameter, c_s is the CO_2 partial pressure at the leaf surface, e_s is the vapor pressure at the leaf surface, e_i is the saturation vapor pressure inside the leaf at the vegetation temperature, P_{atm} is the atmospheric pressure, and b = 2000 is the minimum stomatal conductance when $A_n = 0$. In summary, the equation is useful in that the stomatal conductance (inverse of stomatal resistance) depends simply on carbon dioxide concentration, relative humidity, and two vegetation-dependent parameters.

The above model represents a single leaf. There exist two types of leaves that are assumed to exist at the top layer called the canopy: sunlit leaves and shaded leaves. The determination of type of leaf at the canopy is based on the amount of PAR absorbed. The canopy itself can have an associated gross photosynthetic rate photosynthesis and stomatal conductance which are respectively defined as $A_{sun}L_{sun} + A_{sha}L_{sha}$ where L_{sun} and L_{sha} are leaf area indices and $\frac{1}{r_s^{sun}}L_{sun} + \frac{1}{r_s^{sha}}L_{sha}$. Both of these were found by integrating the original equations over the depth of the canopy [6, 13, 14].

Chapter 4

Sensitivity Analysis: Background and Local-based Methods

Sensitivity analysis is an important consideration when implementing a mathematical model. A mathematical model of a natural process typically will have parameters that represent quantities pertaining to nature. For instance, a model of free falling of an object near the Earth's surface in a vacuum will include the gravitational acceleration parameter $g \approx 9.81 \frac{m}{s^2}$. Noting the approximation, one can see that this parameter has some uncertainty since it was experimentally determined and there exists random error in experimental measurements for example.

A model that contains many parameters can become less useful due to the fact that each parameter contains uncertainties. Many of the parameters may be challenging to measure by experiments. Additionally, some parameters may be unknown. Reducing the number of parameters thus is one appealing option when improving a model. In general, the solution to the problem with uncertainties is the understanding on how uncertainties of the parameter inputs affect the uncertainties of the model output [15]. The method used to perform such a task is sensitivity analysis. More precisely, sensitivity analysis is defined by Saltelli (2004) as "the study of how the uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input" [16]. Figure 5 illustrates how sensitivity analysis is used in the context of a mathematical model.

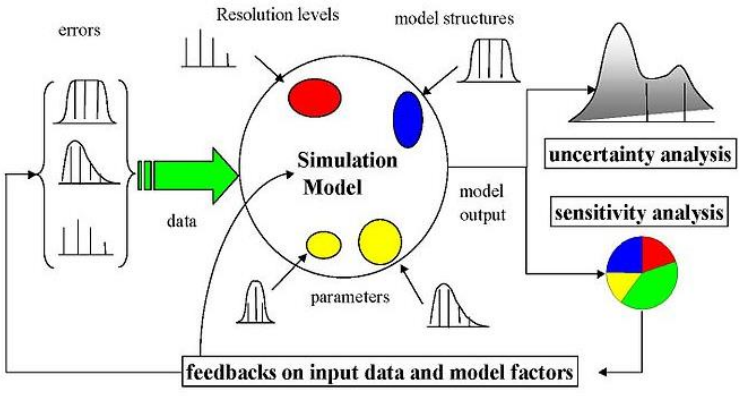


Figure 5. Overall schematic of sensitivity analysis [17]

Uncertainties in the parameters can be due to measurement error, variability found in nature, and the absence of a method to measure the parameters [16]. In this work, the mathematical model is the leaf photosynthesis-stomatal resistance subroutine which has input parameters pertaining to quantities such as vapor pressure inside the leaf and vegetation temperature. A list of these parameters will be given in Chapter 6. All of these parameters will contain uncertainties and thus conducting a sensitivity analysis on these parameters will be appropriate.

The basic idea behind sensitivity analysis is to vary the input parameters, run the model, observe output differences, and repeat the process. For example, if a one input parameter changes by a small amount and the model output changes by a large amount then this parameter can be considered to have high sensitivity. However, if another input parameter is changed by a large amount and the model output changes by a small amount then that parameter is quite insensitive. The model with respect to the insensitive parameter is known to be more robust. This kind of information will allow researchers to perhaps exclude insensitive parameters in improving mathematical models. The goal of this work in the spirit is to utilize sensitivity analysis on the leaf photosynthesis-stomatal resistance subroutine to seek a ranking of important parameters.

The history of sensitivity analysis was influenced by developments in the field of statistics. Gregor Mendel and his work with pea plants and Gosset and his work with Irish hop crops both involved precursor techniques to sensitivity analysis [2]. Throughout time, different approaches to sensitivity analysis were formulated using various frameworks. Sensitivity analysis at the beginning was usually local based while more recently global-based sensitivity analysis has become more widespread [18].

Local sensitivity analysis as the name implies describes sensitivity around a neighborhood of an input point. This will mean that one input factor will vary around a reference value while all the other input factors are kept constant. Mathematically, local sensitivity analysis approaches will generally use partial differentiation to compute sensitivities. However, one drawback to local sensitivity analysis is

that sometimes the derivatives cannot be calculated. Furthermore, another drawback is that the sensitivities calculated pertain to when the input parameter changes by small amount around the reference value [19].

If the model is relatively simple and mathematically well behaved enough a continuous derivative, then the local sensitivity analysis can be obtained by analytic sensitivity functions. Usually these functions are partial derivatives and thus can be written as a function of all other parameters. The other type of sensitivity function is empirical sensitivity functions which involve evaluating at a point for a given parameter while fixing other parameter values. The latter case is going to be used in more complex computer models since analytical functions will be more difficult to derive [2]. The partial derivative of a function will now be discussed to understand why it is used as a sensitivity measure. The mathematics is borrowed from [20].

Consider a function $f(x, y)$, that is a function of only two variable x and y. When the variable y is fixed at $y = b$, then the function $f(x, b)$ is essentially a function of one variable x and denoted as $g(x)$. Therefore, $g(x) = f(x, b)$. The definition of the partial derivative of f with respect to x at (a, b) is:

$$\frac{\partial f}{\partial x}(a, b) = \lim_{h \to 0} \frac{g(a + h) - g(a)}{h} = \lim_{h \to 0} \frac{f(a + h, b) - f(a, b)}{h}$$

Similarly, the partial derivative of f with respect to y at (a, b) is:

$$\frac{\partial f}{\partial y}(a, b) = \lim_{h \to 0} \frac{f(a, b + h) - f(a, b)}{h}$$

In a more general case, for function $f(x_1, x_2, \dots, x_i, \dots, x_{n-1}, x_n)$ of n variables, the partial derivative with respect to the i^{th} variable is:

$$\frac{\partial f}{\partial x_i} = \lim_{h \to 0} \frac{f(x_1, x_2, \dots, x_{i-1}, x_i + h, x_{i+1}, \dots, x_{n-1}, x_n) - f(x_1, x_2, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_{n-1}, x_n)}{h}$$

Therefore, it can be seen that partial derivatives can give information about the rates of change of a function. More exactly, $\frac{\partial f}{\partial x}(x, y)$would mean the rate of change of f with respect to x when y is fixed. This rate can be evaluated at any points (a, b) in the domain. Therefore, the partial derivative quantity is a measure of sensitivity as desired. In matter of fact, $S_p = \frac{\partial f}{\partial p}$ is denoted as the absolute sensitivity for function f with respect to parameter p [2].

Now, consider the function $f(x, y) = Ax + By$. The partial derivatives and thus sensitivities of $f(x, y)$ will only depend on the coefficient values A and B. Then, consider the more complicated function $f(x, y) = Ax^2 + By^2 + Cxy + Dx + Ey + F$. To begin the local sensitivity analysis, an initial point must be chosen. Let the initial point be $(1,1)$. The sensitivities are:

$$S_A = x^2, S_B = y^2, S_C = xy, S_D = x, S_E = y, S_F = 1$$

$$S_A = 1, S_B = 1, S_C = 1, S_D = 1, S_E = 1, S_F = 1$$

Therefore, at the point $(1,1)$, the absolute sensitivities are the same across all six parameters. The significance would be that increasing the parameters by unit 1 will increase the output by unit 1. Additionally, one can take higher order partial derivatives to measure the interaction between parameters. For example, if function f is continuous then the mixed second-order partial derivative $\frac{\partial^2 f}{\partial A \partial x}(x, y) = \frac{\partial^2 f}{\partial x \partial A}(x, y) = 2x$ which then measures the sensitivity when the parameter A and variable x are both changed at the same time. These interaction terms are quite significant since higher order terms parameters may result in higher sensitivities as shown by the example above involving A and x.

Sometimes it is better to normalize the absolute sensitivities so that it is possible to compare all sensitivities. The resulting quantity is called a relative sensitivity, $\bar{S}_p = \left(\frac{\partial f}{\partial p}\right)\left(\frac{p_0}{f_0}\right)$ where p_0 is the normal value of the parameter in question and f_0 is the normal value of the function. This quantity can be derived intuitively by considering changes in percentages instead of actual values. Therefore,

$\left(\frac{\Delta f}{f_0}\right) \div \left(\frac{\Delta p}{p_0}\right) = \left(\frac{\Delta f}{f_0}\right) \times \left(\frac{p_0}{\Delta p}\right) = \left(\frac{\partial f}{\partial p}\right) \left(\frac{p_0}{f_0}\right)$. The relative sensitivity thus can provide information about which parameters will produce the most change in the output based on percent change in the parameters while absolute sensitivity would be based on a fixed size change.

In scenarios when the derivatives are difficult to calculate, then numerical estimation must be used. Numerical estimation based on Taylor Polynomial expansion can be used.

Recall Taylor series expansion of the function f around the point x:

$$f(x+h) = f(x) + h\,\frac{\partial f}{\partial x}(x) + \frac{h^2}{2!}\frac{d^2 f}{dx^2}(x) + \cdots + \frac{h^n}{n!}\frac{d^n f}{dx^n}(x) + \cdots$$

Therefore,

$f(x+h) - f(x) = h\,\frac{df}{dx}(x) + \frac{h^2}{2!}\frac{d^2 f}{dx^2}(\theta)$, where θ is between x and $x+h$.

If h and $\frac{d^2 f}{dx^2}$ are small, then $\frac{df}{dx}(\theta) \approx \frac{f(x+h)-f(x)}{h}$. Otherwise, the second derivative must be accounted for. The Taylor expansion for higher dimensions can be done for multivariable functions and a similar numerical estimation can be done for the partial derivatives [2].

Chapter 5

Sensitivity Analysis: Global-based Methods

More recently, the method of global-based sensitivity analysis is more widely used because they do not have the same limitations as local-based sensitivity analysis approaches. Global-based sensitivity analysis takes into account of the whole parameter space, nonlinearity, interactions between parameters, and models without analytical forms. However, global-based approaches require higher computation costs as a limitation [18]. Global-based sensitivity analysis consists of various techniques such probabilistic variance decompositions, regression analysis, and elementary effects tests [19]. The variance-based sensitivity analysis framework will now be discussed because this type of sensitivity analysis was used with the leaf photosynthesis-stomatal resistance model. The mathematics including some definitions and proofs involving probability is borrowed from [21].

The variance is a mathematical quantity that is defined in a couple of ways. In particular, sensitivity can be defined through variance since sensitivity analysis is about analyzing how the variance of an output is affected by the variance of an input parameter.

It is first important to generalize a mathematical model such as the leaf photosynthesis-stomatal resistance model. In this case, let the model be denoted as $f: \mathbb{R}^n \to \mathbb{R}$ and $X \to Y = f(X)$ where Y is the output, $X = (X_1, \dots, X_n)$ are n independent inputs (in this case, input parameters). The function f does not have to have an analytical form nor is numerical estimation required.

First, set X_i equal to its true value denoted as x_i^*. Then, one way to intuitively define the importance of an input variable X_i is to see how it affects the variance of Y. Mathematically, this quantity can be written a $Var(Y \mid X_i = x_i^*)$ which is known as the conditional variance of Y given $X_i = x_i^*$. However, one problem with this definition is that frequently the true value x_i^* of X_i is unknown. The correction to this problem is to average of the conditional variances under all possible values for x_i^*. This

average is written as $E[Var(Y|X_i = x_i^*)]$ which is known as the expected value of the conditional variances of Y given $X_i = x_i^*$ [22].

The expected value is an important quantity used in mathematics and statistics. The most general definition of an expected value is the value of a random variable X is called the mean or the weighted average of X. More precisely, the expected value of a continuous random variable X with the probability density function f is defined by:

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} \mathrm{xf(x)\,dx}$$

A geometric analog to expected value is to think of a graph of the function $f(X)$ that has been cut out of cardboard with uniform density. The expected value of X is the point on the x-axis at which the graph will balance on a given edge orthogonal to the x-axis [21].

However, random variables are not constantly equal to its expected value. In nature, random variables vary from their expected values all the time. These fluctuations from the expected values can be quantified by averaging the magnitude of such fluctuations. This quantity is then called the variance. Precisely, the variance of a continuous random variable X with the probability density function f is defined by:

$$Var(X) = E\left[\big(X - E(X)\big)^2\right] = \int_{-\infty}^{\infty} \big(x - E(X)\big)^2 \mathrm{f(x)\,dx}$$

In probability, there arises a situation when one needs to calculate conditional probabilities. The conditional probability of A given B is the probability that A occurs given that B has already occurred. More precisely, the conditional probability of A given B is denoted by $P(A|B) = \frac{\mathrm{P(AB)}}{\mathrm{P(B)}}$ where $P(AB)$ is the probability that A and B happen (the joint occurrence) and $P(B) > 0$ [21].

Now, consider

$$Var(X) = E\left[\left(X - E(X)\right)^2\right]$$

$$= E(X^2 - 2 * E(X) * X + E(X)^2)$$

$$= E(X^2) - 2 * E(X) * E(X) + E(X)^2$$

$$= E(X^2) - 2 * E(X)^2 + E(X)^2$$

$$= E(X^2) - E(X)^2$$

Therefore, $Var(X) = E(X^2) - E(X)^2$. Furthermore, by similar reasoning, $Var(X|Y) = E(X^2|Y) - E(X|Y)^2$. An important quantity to examine is then $E(X|Y)$. Observe the following:

$$E[E(X|Y)] = \int_{-\infty}^{\infty} E(X|Y = y) f_Y(y) dy$$

$$= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x f_{X|Y}(x|y) \, \mathrm{dx} \right) f_Y(y) \, dy$$

$$= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y) dy \right) \mathrm{dx}$$

$$= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} \frac{f(x,y)}{f_Y(y)} f_Y(y) dy \right) \mathrm{dx}$$

$$= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f(x,y) dy \right) \mathrm{dx}$$

$$= \int_{-\infty}^{\infty} x f_X(X) \, \mathrm{dx}$$

$$= E(X)$$

Therefore, $E[E(X|Y)] = E(X)$. One can then conclude that the expected value of X is the weighted average of conditional expected values of X given Y over all possible values of Y [21].

Finally, consider

$$Var(X|Y) = E(X^2|Y) - E(X|Y)^2$$

which implies $E[Var(X|Y)] = E[E(X^2|Y)] - E[E(X|Y)^2] = E(X^2) - E[E(X|Y)^2]$.

Additionally,

$$Var(E[X|Y]) = E[E(X|Y)^2] - (E[E(X|Y)])^2 = E[E(X|Y)^2] - [E(X)]^2.$$

Therefore,

$Var(X) = E[Var(X|Y)] + Var(E[X|Y])$. Similarly, $Var(Y) = E[Var(Y|X)] + Var(E[Y|X])$.

Now from the original motivation of defining sensitivity from variance,

$$Var(Y) = E[Var(Y|X_i = x_i^*)] + Var(E[Y|X_i = x_i^*]) .$$

Recall that the intuitive definition of quantifying the importance of input X_i is $E[Var(Y|X_i = x_i^*)]$. For a constant $Var(Y)$, a lower $E[Var(Y|X_i = x_i^*)]$ would imply a higher $Var(E[Y|X_i = x_i^*])$.

The quantity $Var(E[Y|X_i = x_i^*])$, named as the variance of the conditional expected value (also, written as $Var(E[Y|X_i])$) can be thought of the importance of $X_i$on the variance of Y. Therefore, $Var(E[Y|X_i = x_i^*])$ is considered to be the sensitivity of Y to X_i. It is to be noted that the more important is X_i, the greater is $Var(E[Y|X_i])$. Finally, $Var(E[Y|X_i])$ is normalized to $\frac{Var(E[Y|X_i])}{Var(Y)}$ so that it can take a value between 0 and 1. $\frac{Var(E[Y|X_i])}{Var(Y)}$ is known as the first-order sensitivity index, correlation ratio, and importance measure, which quantifies how input X_i mainly affects output Y [22].

As was seen in the local-based approach, interactions between parameters can be measured. This is accomplished by variance decomposition. The variance of Y can be decomposed by the following way:

$$Var(Y) = \sum_{i=1}^{n} V_i + \sum_{1 \le i < j \le n} V_{ij} + \cdots + V_{1\ldots n}$$

Where

$V_i = Var(E[Y|X_i])$

$V_{ij} = Var\big(E\big[Y\big|X_i, X_j\big] - E[Y|X_i] - E\big[Y\big|X_j\big]\big)$

$V_{ijk} = Var\big(E\big[Y\big|X_i, X_j, X_k\big] - E\big[Y\big|X_i, X_j\big] - E[Y|X_i, X_k] - E\big[Y\big|X_j, X_k\big] + E[Y|X_i] + E\big[Y\big|X_j\big] + E[Y|X_k]\big)$

Etc.

Because the X are independent inputs,

$V_i = Var(E[Y|X_i])$

$V_{ij} = Var\big(E\big[Y\big|X_i, X_j\big]\big) - V_i - V_j$

$V_{ijk} = Var\big(E\big[Y\big|X_i, X_j, X_k\big]\big) - V_{ij} - V_{ik} - V_{jk} - V_i - V_j - V_k$

Etc.

The first-order sensitivity index is $S_i = \frac{V_i}{Var(Y)}$ which is consistent with the previous definition. The second-order sensitivity index is then $S_{ij} = \frac{V_{ij}}{Var(Y)}$ which quantifies interactions between inputs X_i and X_j. The third-order sensitivity index is then $S_{ijk} = \frac{V_{ijk}}{Var(Y)}$ which quantifies interactions between inputs X_i, X_j, and X_k. Higher order indices are defined in a similar manner. The number of all order indices can be calculated by the formula $2^n - 1$ where n is the number of inputs. Finally, the total sensitivity index S_{T_i} can be computed by summing all order indices relating to X_i. For example, $S_{T_1} = S_1 + S_{12} + S_{12} + S_{123}$ for the input X_1 and three total inputs, X_1, X_2, and X_3 [22].

Besides using a variance-based approach, regression analysis can be used as long as the model is linear or close to being linear. This is more limited than variance-based approaches which can handle cases where there is non-monotonicity, non-linearity, and non-additivity. Regression analysis is based on fitting a linear model to data. The general linear model is $f(\boldsymbol{x}) = \boldsymbol{y} = b_0 + \sum_{i=1}^{n} b_i x_i + \varepsilon$ where the output vector is $\boldsymbol{y}$, number of parameters is n, i^{th} parameter is x_i, estimated coefficient for x_i is b_i, and ε is the random error. The b_i's can be calculated by the least squares method and minimizing ε. Afterwards, the quantity known as the standardized regression coefficient (SRC) can be defined by $SRC(x_i) = \beta_i = b_i \frac{\hat{s}_j}{\hat{s}}$ where $\hat{s}_j$ is the standard deviation for input parameters and $\hat{s}$ is standard deviation for outputs. This standardization of the b_i's is meaningful since it allows for changing an input parameter by a fixed fraction of its standard deviation while all other parameters are fixed. Because the b_i's are essentially slopes, $\beta_i' s$ measure sensitivities [23]. There are other sensitivity measures from regression analysis such as PCC (Partial Correlation Coefficient) and PEAR (Pearson Product Moment

Correlation Coefficient). When the model is not linear, a ranked transformed statistic can be computed known as the SRRC (Standardized Rank Regression Coefficients) [24]. SRRC involves first ranking the outputs from largest output to smallest output and then performing regression analysis on the ranking. Thus, the limitation is that the model is transformed into one based on the rankings. This makes SRRC more of a qualitative measure of sensitivity than a quantitative one. Another limitation is that monotonicity is preferred when using the model [25].

The last type of global-based sensitivity analysis approach is an elementary effect test which is exemplified by the Morris Method, a screening method. It has the advantage since it requires a relatively much lower computation cost. The limitation is that that the sensitivity measures are qualitative rather than quantitative. Thus, a ranking can be made but measuring how much one parameter is more sensitive than the other is difficult [24]. In the Morris Method, the ranges are divided into (p-1) intervals of equal sizes. A quantity called the elementary effect is calculated by the following: $EE_i = \frac{1}{\tau_y}\frac{f(x_1*,\ldots,x_i*+\Delta,\ldots x_n*)-f(x_1*,\ldots,x_i*,\ldots x_n*)}{\Delta}$, where $\Delta = \frac{p}{2(p-1)}$. The elementary effects' mean and standard deviation are calculated. The absolute value of the mean will represent the overall importance of a parameter while the standard deviation will represent nonlinear and interaction effects [18]. The sensitivity measures can either be the absolute value of the mean or the standard deviation and then be ranked accordingly [16].

Chapter 6

Materials and Procedure

The leaf photosynthesis and stomatal resistance subroutine was used in the form of a "stomata" functional test with a main user interface provided by Dr. Dali Wang and his team from the Climate Change Science Institute of Oak Ridge National Laboratory [26]. Figure 6 is a snapshot of the interface.

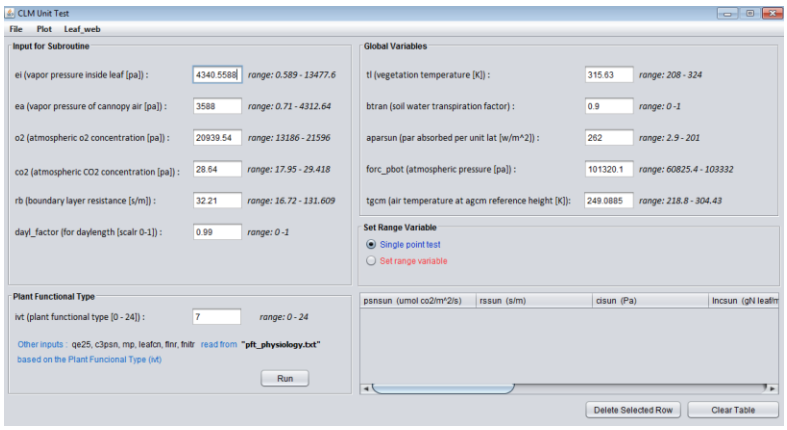


Figure 6. "Stomata" functional test main user interface

The subroutine contains 11 input parameters, 1 plant functional type parameter (categorical), and 5 outputs. Table 1 lists all the parameters used in the subroutine with given ranges:

Abbreviation	Entire Name	Units	Range
Input for Subroutine			
ei	vapor pressure inside leaf	Pa	0.589 - 13477.6
ea	vapor pressure of canopy air	Pa	0.71 - 4312.64
o2	atmospheric O_2 concentration	Pa	13186 - 21596
co2	atmospheric CO_2 concentration	Pa	17.95 - 29.418
rb	boundary layer resistance	s/m	16.72 - 131.609
dayl_factor	daylength	scalar	0 - 1
Global Variables			
tl	vegetation temperature	K	208 - 324
btran	soil water transpiration factor	scalar	0 - 1
aparsun	par* absorbed per unit sunlit lai**	W/m^2	2.9 - 201
forc_pbot	atmospheric pressure	Pa	60825.4 - 103332
tgcm	air temperature at agcm*** reference height	K	218.8 - 304.43
Plant Functional Type			
ivt	plant function type	scalar	0 - 24 (discrete)
Output for Subroutine			
psnsun	sunlit leaf photosynthesis	μmol $CO_2/m^2/s$	
rssun	sunlit stomatal resistance	s/m	
cisun	sunlit intracellular CO_2	Pa	
lncsun	leaf N concentration per unit projected sunlit lai**	gN $leaf/m^2$	
vcmxsun	sunlit maximum rate of carboxylation	μmol $CO_2/m^2/s$	
*photosynthetically active radiation			
**leaf area index			
***atmospheric general circulation model			

Table 1. List of all parameters for the leaf photosynthesis-stomatal resistance subroutine

The psnsun refers to the leaf photosynthetic rate and the rssun refers to stomatal resistance as mention in Chapter 3. The plant functional types are listed in Table 2:

PFT	Name
1	Needleleaf evergreen tree - temperate
2	Needleleaf evergreen tree - boreal
3	Needleleaf deciduous tree - boreal
4	Broadleaf evergreen tree - tropical
5	Broadleaf evergreen tree - temperate
6	Broadleaf deciduous tree - tropical
7	Broadleaf deciduous tree - temperate
8	Broadleaf deciduous tree - boreal
9	Broadleaf evergreen shrub - temperate
10	Broadleaf deciduous shrub - temperate
11	Broadleaf deciduous shrube - boreal
12	C_3 arctic grass
13	C_3 grass
14	C_4 grass
15	$Crop_1$
16	$Crop_2$

Table 2. List of plant functional types [14]

The sensitivity analysis methods were implemented using a software called SimLab (Simulation Laboratory for Uncertainty and Sensitivity Analysis) which was developed the Joint Research Centre of the European Commission. Its current version 2.2 can be downloaded at [27]. SimLab is a free professional tool that has been continuously improved since 1985. Knowing how to use SimLab was very important since it can provide many features that are relevant to the study of sensitivity analysis of a model. SimLab is designed for global-based sensitivity analysis which was covered in Chapter 5. Also, Simlab is made of three modules which are used in succession. The first model is known as the Statistical Pre-Processor module, the second is the Model Execution module, and the third is the Statistical Post-Processor module. The three modules can be seen in Figure 7.

The function of the Statistical Pre-Processor module is to generate sample data points in the space of the input parameters. The function of the Model Execution module is to run the model for each sample point created from the Statisical Pre-Processor module. Lastly, the function of the Statistical Post-Processor module is to execute the uncertainty and sensitivity analysis. The first module involves choosing probability distribution functions for the uncertainty of each input factor. Because the

probability distributions and statistical properties for each input factor were unknown, it was decided to use the uniform distribution for each input factor based on the ranges given in Table 1.

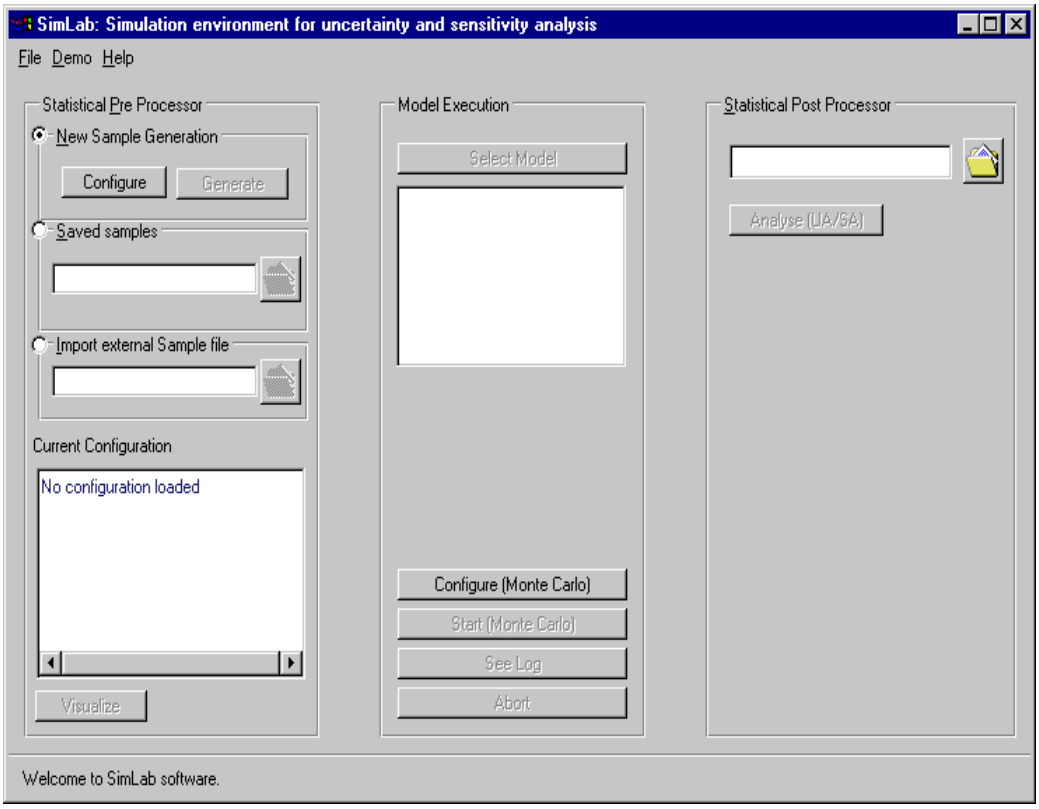


Figure 7. Interface of SimLab with its 3 modules [25]

From the mathematics presented in [21], a random variable X is said to be uniformly distributed over an interval (a,b) if its probability density function is given by:

$$f(x) = \begin{cases} c, & if\ a < x < b \\ 0, & elsewhere \end{cases}$$

Because $f(x)$ is a probability density function, $\int_b^a f(x)\,dx = 1$ which implies that $\int_b^a c\,dx = c(b-a) = 1$. Therefore, $c = \frac{1}{b-a}$.

Since the analysis at this point in the research is mainly exploratory, it is justified to use uniform distribution. There is also no underlying data that could be used. Even if there were underlying data, there is a cost for getting better probability distribution functions.

After the selection of uniform distribution and setting the ranges, the next step is to choose sampling methods. The sampling methods that can be chosen in SimLab are FAST, Fixed samples, Latin Hypercube, Morris, QuasiRandom LpTau, Random, and Sobol. The sampling method is mainly chosen based on the input factors. This research seeks to utilize various methods to see how they affect sensitivity analysis. The sampling methods used in this work were Latin Hypercube, Extended Fast, Sobol, Random, and Morris.

The sampling method is one aspect of the Monte Carlo simulation used in generating input data. The idea is that the model is evaluated many times through random numbers. A probability distribution has to be chosen for each parameter characterizing uncertainty. If the probability distribution is not known, then a uniform distribution can be chosen as was done in this work [15]. Besides the probability distributions, the sample size has to be assigned for each parameter. A large enough evaluation has to be chosen or else the sample is not a good representation of the population. The samples themselves are created by selecting points by probability distribution function. The input data is evaluated into the model which returns a set of outputs of same sample size. Uncertainty can then be quantified or visually represented on a graph.

Latin Hypercube Sampling (LHS) is a type of Monte Carlo sampling method. From LHS, one can generate an estimation of average model outputs. Compared to Random Sampling, LHS has the same accuracy but through input. The basis of LHS is stratified sampling without replacement. The distribution is divided into N equal intervals where N is the sample size. Each interval is then sampled once. Each factor thus contains N realizations from the N intervals. A random realization is chosen from each factor and this is considered to the first input of the sample [25]. Therefore, this guarantees that the sample is

able to represent values from all over the distribution (high, low, and middle). For instance, if N = 4, then the range [0, 1] the LHS would contain values from the ranges [0, 0.25], [0.25, 0.50], [0.50, 0.75], and [0.75, 1] rather than randomly having values just from [0, 0.25] for example.

Random Sampling is another Monte Carlo sampling method but it has less accuracy in terms of generating means and the probability distributions when compared to LHS [25]. In Random Sampling, the points are selected from joint distribution of the input parameters assuming they are not independent. If the parameters are independent, then the marginal distributions are used. The Random Distribution is known as pseudo-random since the random numbers are generated by a deterministic process and an initial value known as a seed. The joint distribution between uniform random variables can be understood by the following:

Let $X_1, X_2, \dots X_n$ be n random variables with joint probability density function $f(x_1, x_2, \dots, x_n)$. Then the n random variables are independent if and only if $f(x_1, x_2, \dots, x_n) = f_{X_1}(x_1) \times f_{X_2}(x_2) \times \dots \times f_{X_n}(x_n)$ where $f_{X_i}(x_i)$ is the marginal density for random variable X_1. Thus, assuming $X_1, X_2, \dots, X_n$ are n independent uniform random variables with ranges $[a_i, b_i]$ for X_i, then

$$f(x_1, x_2, \dots, x_n) = \begin{cases} c, & if\ \boldsymbol{x} \in A \\ 0, & elsewhere \end{cases}$$

where $c = \frac{1}{\prod_{i=1}^{n}(b_i - a_i)}$ and $A = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_n, b_n]$. In this work, *n* is 11.

There were a minimum of two characterizations for each sampling method which were a seed for a random number generator algorithm that is used and the number of executions (sample size). The seed was recommended to be at least a seven digit number. The seed chosen for all sampling methods was 1234567. The number of executions will generate the cardinality of the sample set. However, choosing the number of executions is less straightforward because each method's accuracy is dependent on the number of executions. The appropriate number of executions for each method was determined based on convergence studies in [23].

The number of executions determined by [23] was 4000 for LHS, 4000 for Random, and 22803 for Extended FAST. In [23], the Sobol method is considered to be the reference method with a number of executions to be 49,152. The authors of [23] had varied the number of executions for various methods and compared the rankings with those produced by the Sobol method. The Morris method was not compared because there are three characterizations instead of two. Besides the seed and number of executions, the number of "levels" is a factor which corresponds to quantiles of the input factors distribution. In this work, level was chosen to be 4 as in [28].

In particular, the actual number of executions chosen in this work was higher and more conservative than in [23] to ensure even more accuracy. Thus, the number of executions was 49152 for Sobol, 24915 for Extended FAST, 120 for Morris, 4500 for Random, and 4500 for LHS.

As for the sensitivity analysis methods themselves, the Sobol and Extended FAST methods are variance-based methods as presented in Chapter 5. In the Sobol Method, the S_i quantity is estimated through variance decompositions and Monte Carlo integration [25]. The inputs must be independent in the Sobol method but it can handle nonlinear models. More information about the Sobol Method can be found in [29].

The FAST method is known as the Fourier Amplitude Sensitivity Test. Both monotonic and non-monotonic models can be used with FAST which is a good advantage. The idea is to transform a multidimensional integral over all inputs that are uncertain into a single one-dimensional integral. This integral is known as a search curve that "scans" all of the parameters. The Fourier series decomposition is utilized to calculate how much each input parameter contributes to variance [25]. The Extended FAST can provide sensitivity indices first order and total order. More information about the Extended FAST method can be found in [31].

The Model Execution module is the module after the Statistical Pre-Processor module. In the Model Execution module, SimLab can perform the Monte Carlo analysis on the input factors sample set

that was created in the previous module. In this module, the user can choose between using an internal model or an external model depending on the design of the experiment. In the internal model, the generated sample is manipulated through mathematical equations such as linear equation that the user can program. However, this work uses a standalone leaf photosynthesis-stomatal resistance subroutine involving more complicated equations as described in Chapter 3. Therefore, the external model option is used to evaluate the generated sample. One option that is used to link SimLab and the external model is through Microsoft Excel. SimLab allows this "Excel switch" option when using the Processing Module. By using this option, SimLab will produce an Excel Worksheet titled "Inputs" with the name of the input factors and the corresponding sample. At this point, SimLab expects the user to input the corresponding outputs into another provided Excel Worksheet titled "Outputs." In this work, the user manually takes the data in the worksheet and adds them to a new worksheet. It is reformatted so that it can be read by the subroutine. The subroutine already has the function to produce Excel spreadsheet of the output data after model evaluation. This output data then is simply transferred back to the "Outputs" Excel Worksheet from SimLab and upon closing, the Model Evaluation module ends.

Finally, the Statistical Post-Processor Module is opened up to the user to perform uncertainty analysis and sensitivity analysis for any combinations of the factors. The Uncertainty Analysis option can provide means, variances, and probability distribution functions estimated from the model predictions. The Sensitivity Analysis option will provide sensitivity measures in the form of tables and graphs. The sensitivity measures and the type of graphs will depend on which method is used. These will be explored in Chapter 7. A summary of the procedure in performing the data collection in this work is depicted in Figure 8.

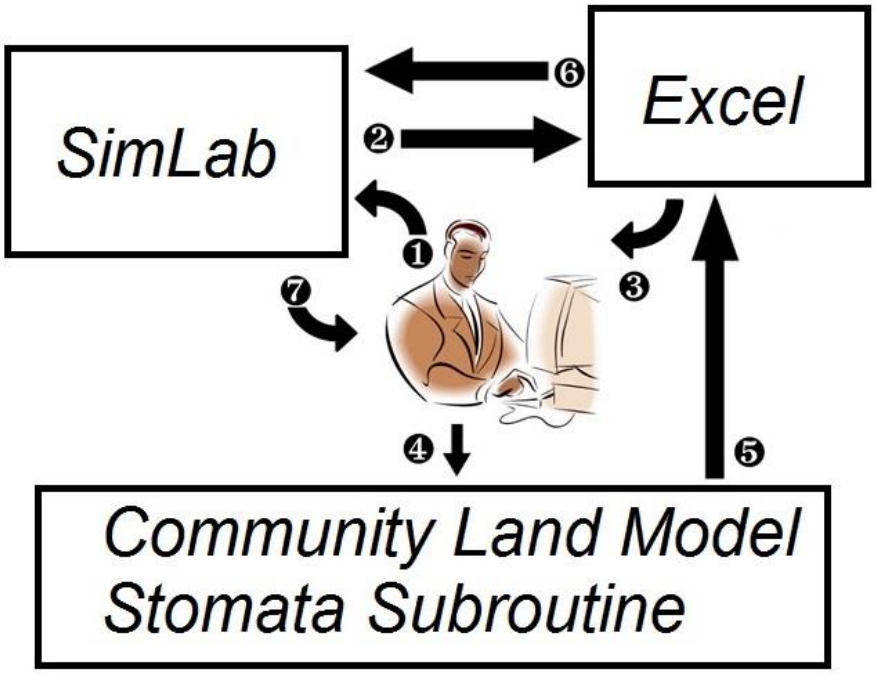


Figure 8. Flow chart of data collection using SimLab and the leaf photosynthesis-stomatal resistance subroutine

Chapter 7

Results: Broadleaf Deciduous Tree - Temperate (PFT = 7)

Five methods of sensitivity analysis were tested on data pertaining to broadleaf deciduous trees in temperate zones. The 5 possible outputs were psnsun, rssun, cisun, lncsun, and vcmxsun. However, the lncsun output was constant regardless of input and sensitivity analysis could not be performed on the lncsun output. Thus, lncsun data is omitted. It was constant in the leaf photosynthesis-stomatal resistance subroutine because nitrogen limitation was not considered.

After running the procedure explained in Chapter 6, measures of sensitivity analysis were obtained in the form of tables and graphs. One example of a table is presented in Table 3.

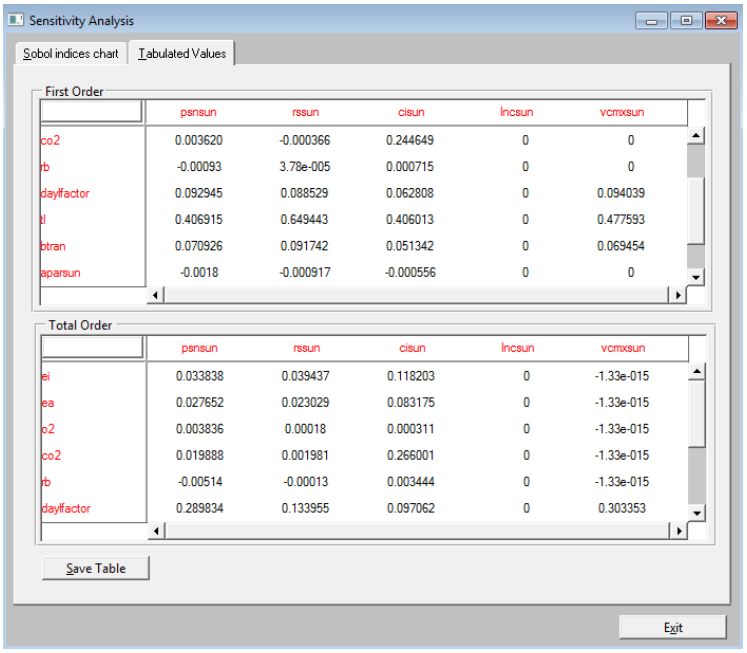


Table 3. Table of sensitivity indices from the Sobol Method

The actual value of sensitivity measures are not the focus of this work. Rather, this work will concentrate more on the relative comparison of sensitivity measures across the different methods, outputs, input parameters, and plant functional types. In this work, the tabulated values were graphed using Excel and then comparisons could be made.

The bar graphs of first order and total order sensitivity indices from the Sobol Method on psnsun, rssun, cisun, and vcmxsun are the following:

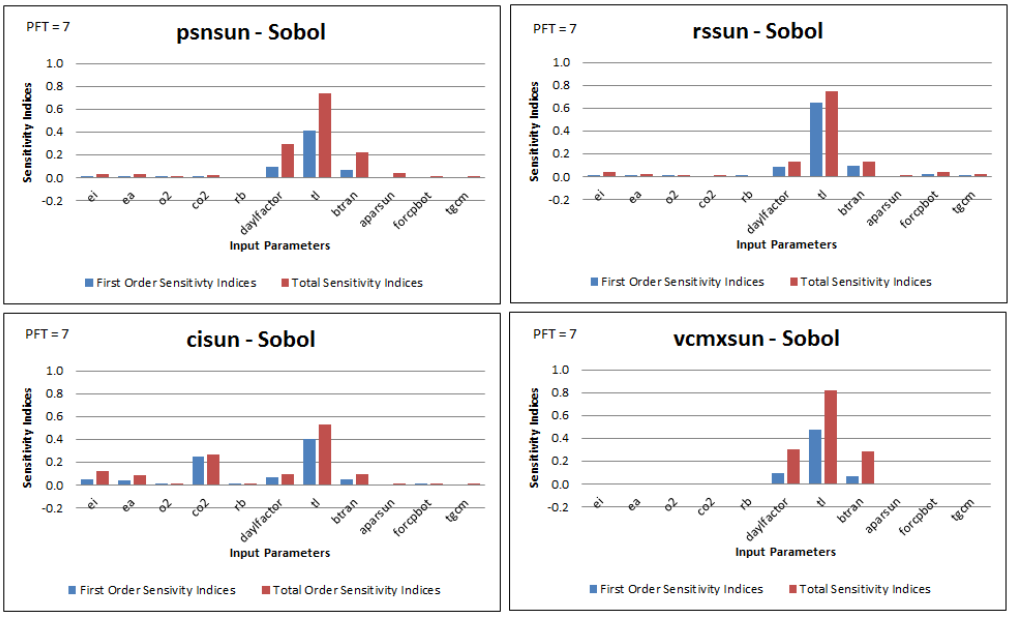


Figure 9. The Sobol Method

The bar graphs of first order and total order sensitivity indices from Extended FAST Method on psnsun, rssun, cisun, and vcmxsun are the following:

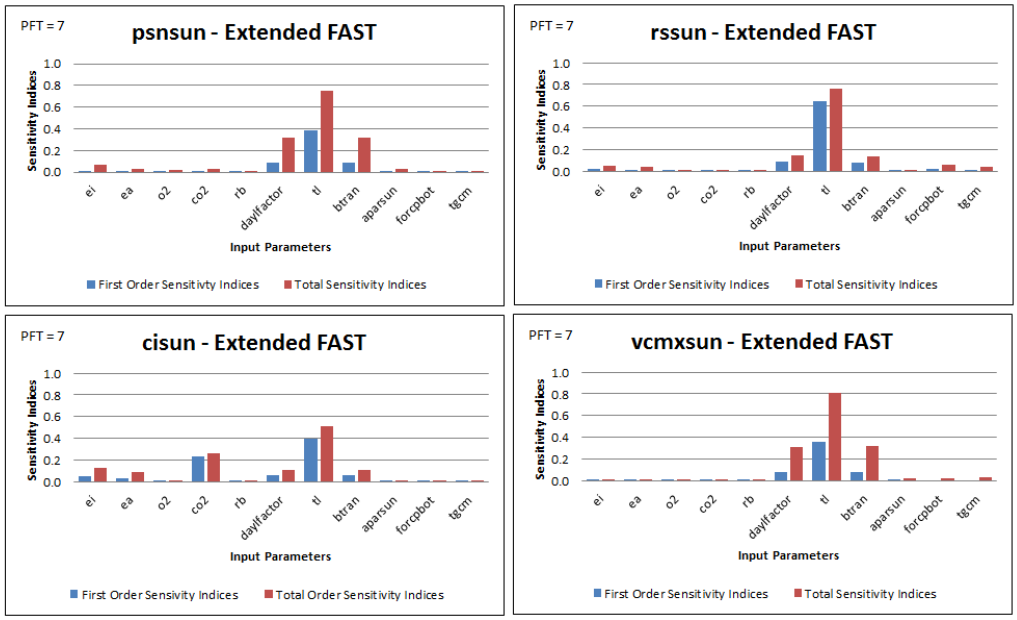


Figure 10. The Extended FAST Method

The bar graphs of sensitivity measures (mean and standard deviations of elementary effects) from the Morris Method on psnsun, rssun, cisun, and vcmxsun are the following:

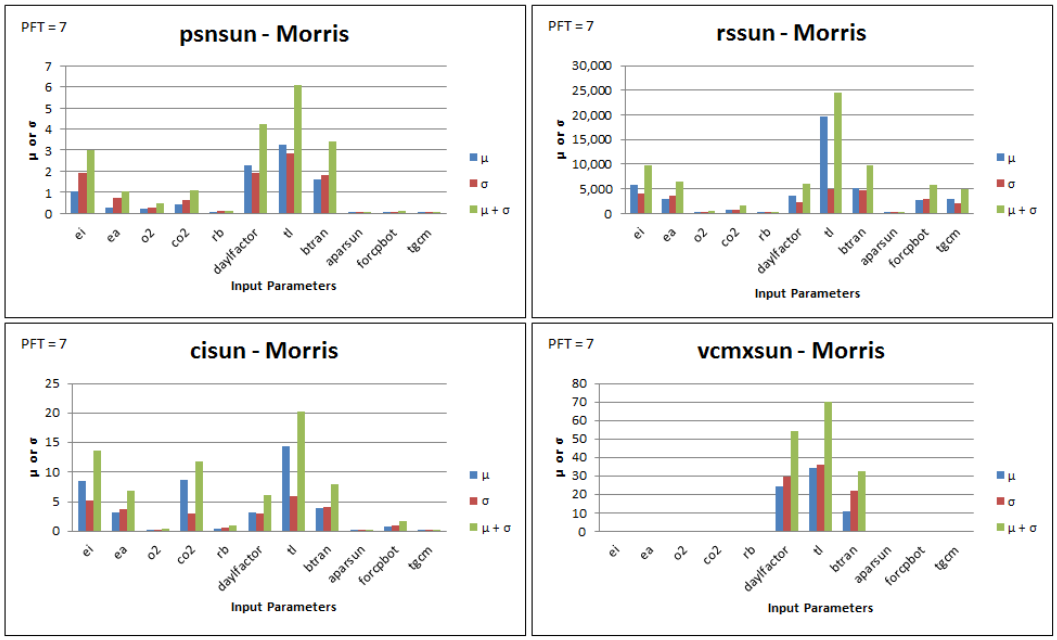


Figure 11. The Morris Method

The bar graph of sensitivity measures (SRRC) from Regression Analysis with LHS Sampling on psnsun, rssun, cisun, and vcmxsun is the following:

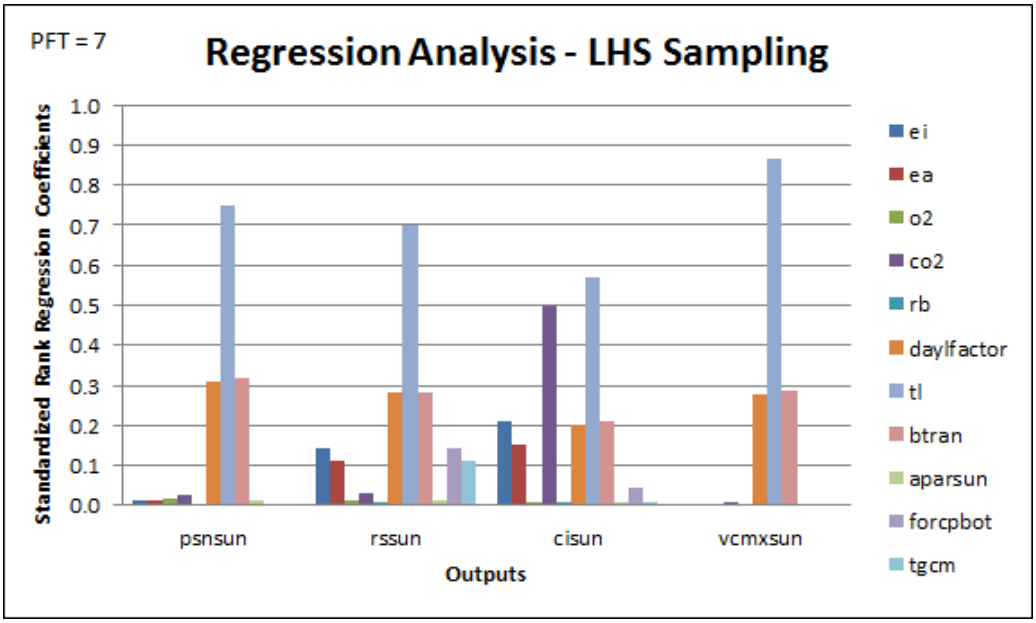


Figure 12. Regression Analysis based on Latin Hypercube Sampling

The bar graph of sensitivity measures (SRRC) from Regression Analysis with Random Sampling on psnsun, rssun, cisun, and vcmxsun is the following

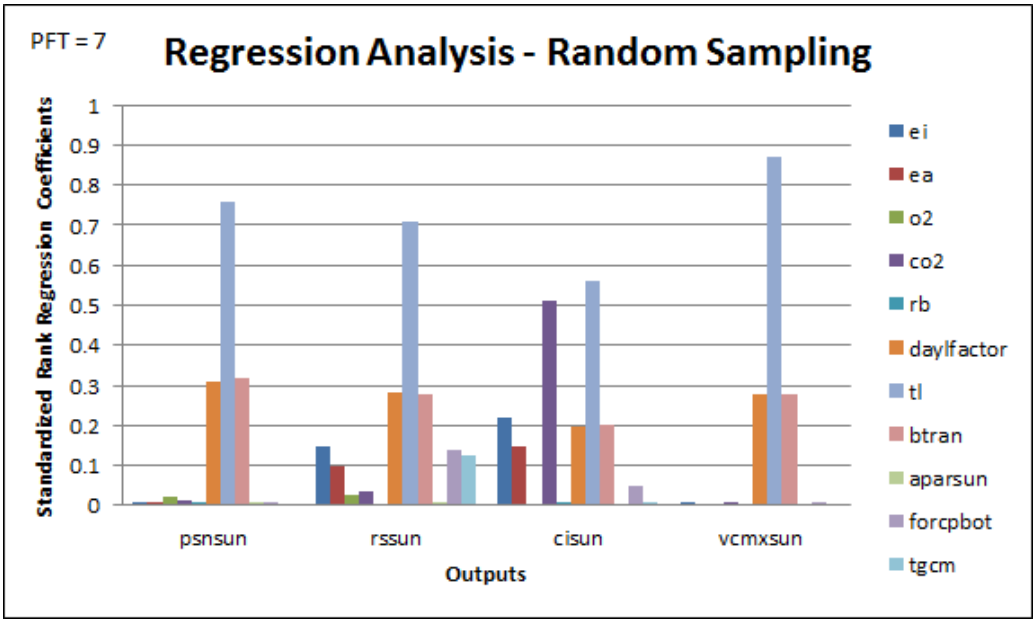


Figure 13. Regression Analysis based on Random Sampling

A summary of the rankings based on the five methods were compiled in the tables below:

Output (PFT = 7)	psnsun				
Method	Sobol	EFAST	Morris	LHS SRRC	Random SRRC
Sensitivity decreases from top to bottom	tl	tl	tl	tl	tl
	daylfactor	btran	daylfactor	btran	btran
	btran	daylfactor	btran	daylfactor	daylfactor
			ei		
			co2		
			ea		

Output (PFT = 7)	rssun				
Method	Sobol	EFAST	Morris	LHS SRRC	Random SRRC
Sensitivity decreases from top to bottom	tl	tl	tl	tl	tl
	daylfactor	daylfactor	btran	btran	daylfactor
	btran	btran	ei	daylfactor	btran
			ea	forcpbot	ei
			daylfactor	ei	forcpbot
			forcpbot	tgcm	tgcm
			tgcm	ea	ea

Output (PFT = 7)	cisun				
Method	Sobol	EFAST	Morris	LHS SRRC	Random SRRC
Sensitivity decreases from top to bottom	tl	tl	tl	tl	tl
	co2	co2	ei	co2	co2
	ei	ei	co2	ei	ei
	daylfactor	btran	btran	btran	btran
	btran	daylfactor	ea	daylfactor	daylfactor
	ea	ea	daylfactor	ea	ea
			forcpbot		

Output (PFT = 7)	vcmxsun				
Method	Sobol	EFAST	Morris	LHS SRRC	Random SRRC
Sensitivity decreases from top to bottom	tl	tl	tl	tl	tl
	daylfactor	btran	daylfactor	btran	btran
	btran	daylfactor	btran	daylfactor	daylfactor

Table 4. Rankings for Broadleaf Deciduous Tree - Temperate

Chapter 8

Results: Needleleaf Evergreen Tree - Temperate (PFT = 1)

The same five methods of sensitivity analysis were tested on data pertaining to needleleaf evergreen trees in temperate zones. The 5 possible outputs were psnsun, rssun, cisun, lncsun, and vcmxsun. Again, the lncsun output was constant regardless of input and sensitivity analysis could not be performed on the lncsun output. Therefore, the lncsun data is omitted.

The bar graphs of first order and total order sensitivity indices from the Sobol Method on psnsun, rssun, cisun, and vcmxsun are the following:

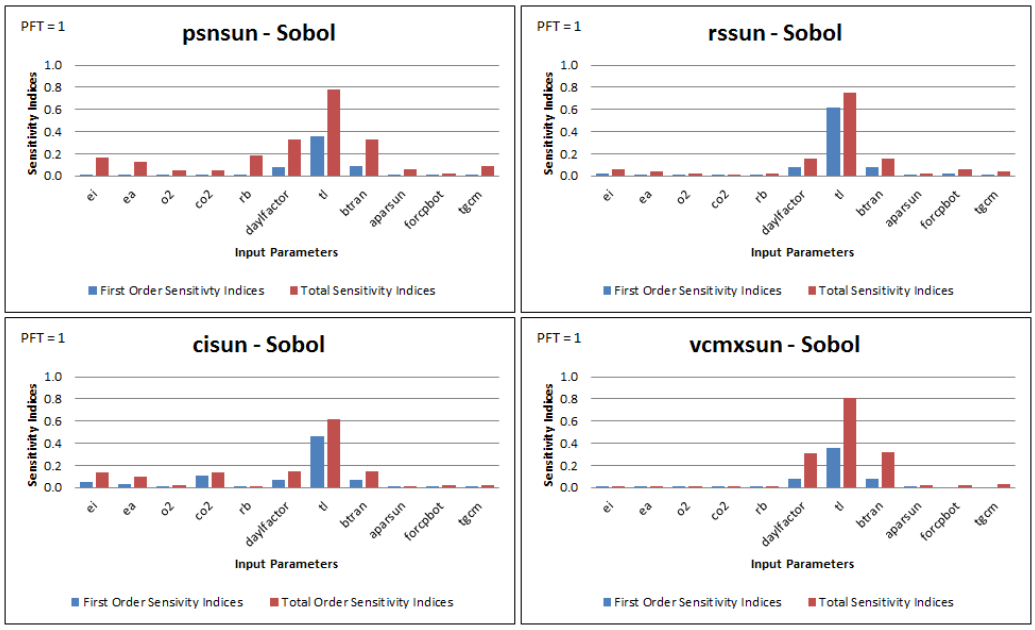


Figure 14. The Sobol Method

The bar graphs of first order and total order sensitivity indices from Extended FAST Method on psnsun, rssun, cisun, and vcmxsun are the following:

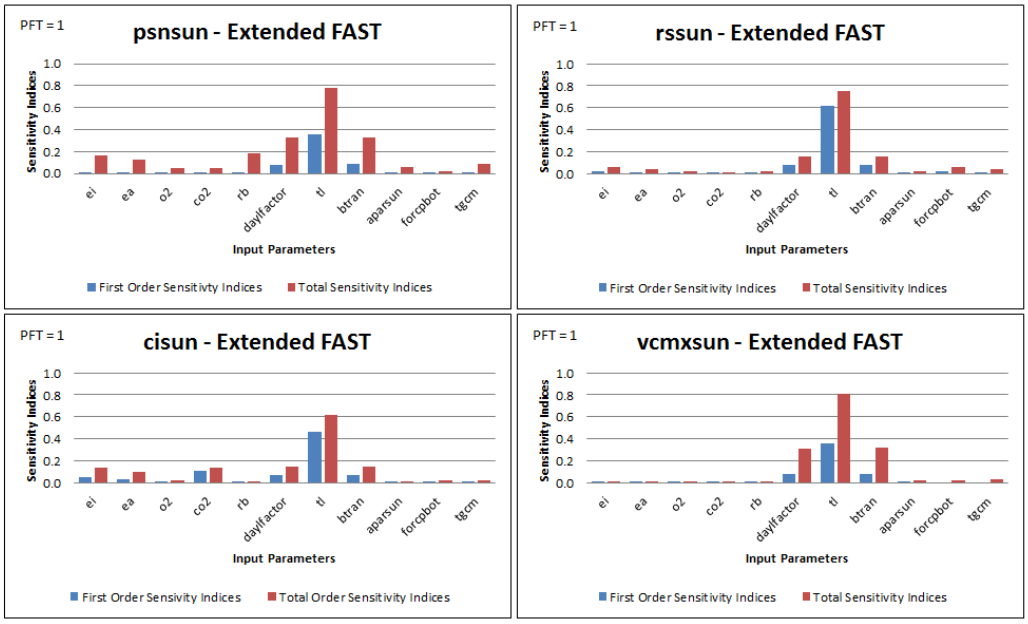


Figure 15. The Extended FAST Method

The bar graphs of sensitivity measures (mean and standard deviations of elementary effects) from the Morris Method on psnsun, rssun, cisun, and vcmxsun are the following:

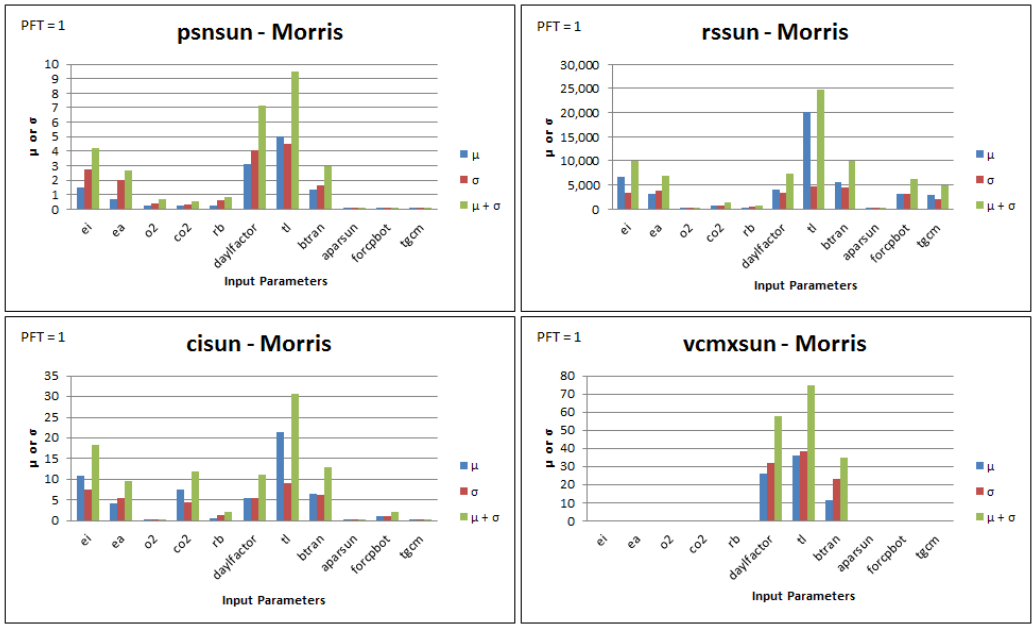


Figure 16. The Morris Method

The bar graph of sensitivity measures (SRRC) from Regression Analysis with LHS Sampling on psnsun, rssun, cisun, and vcmxsun is the following:

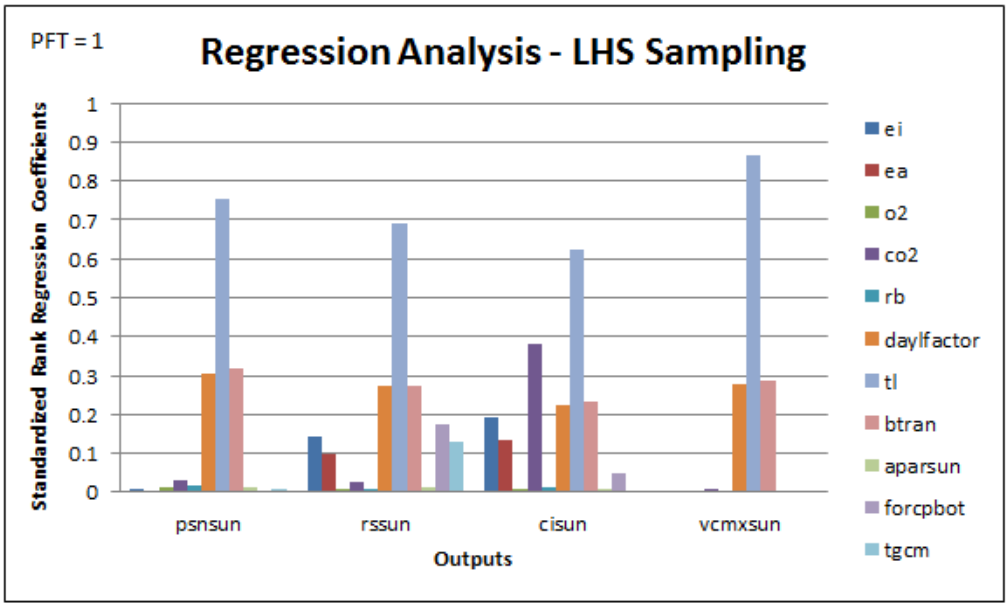


Figure 17. Regression Analysis based on Latin Hypercube Sampling

The bar graph of sensitivity measures (SRRC) from Regression Analysis with LHS Sampling on psnsun, rssun, cisun, and vcmxsun is the following:

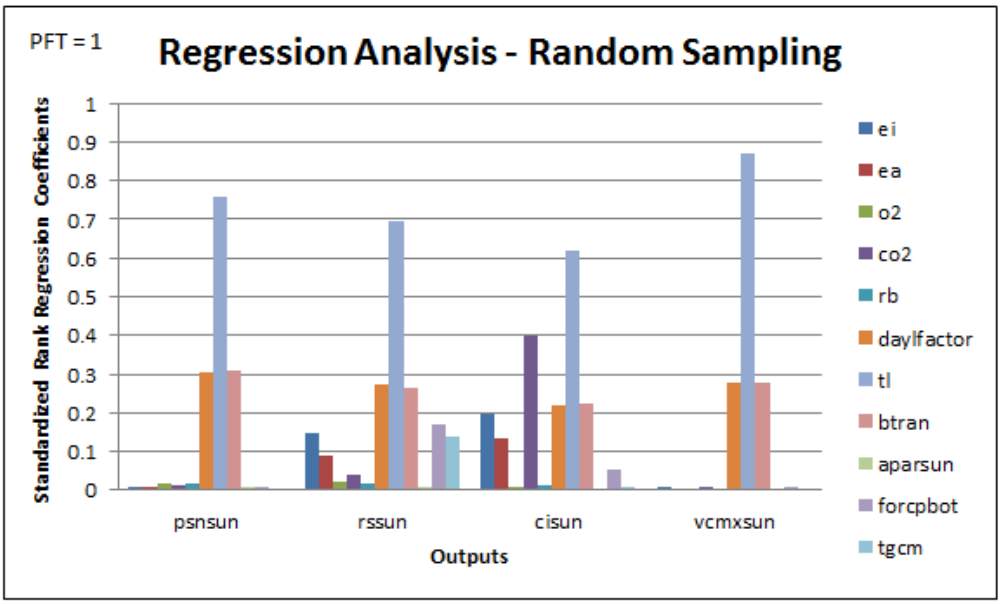


Figure 18. Regression Analysis based on Random Sampling

After producing the graphs and comparing them, a summary of the rankings based on the five methods were compiled in the tables below:

Output (PFT = 1)	psnsun				
Method	Sobol	EFAST	Morris	LHS SRRC	Random SRRC
Sensitivity decreases from top to bottom	tl	tl	tl	tl	tl
	daylfactor	daylfactor	daylfactor	btran	btran
	btran	btran	ei	daylfactor	daylfactor
	rb	rb	btran		
	ei	ei	ea		
	ea	ea			

Output (PFT = 1)	rssun				
Method	Sobol	EFAST	Morris	LHS SRRC	Random SRRC
Sensitivity decreases from top to bottom	tl	tl	tl	tl	tl
	btran	btran	btran	daylfactor	daylfactor
	daylfactor	daylfactor	ei	btran	btran
			daylfactor	forcpbot	forcpbot
			ea	ei	ei
			forcpbot	tgcm	tgcm
			tgcm	ea	ea

Output (PFT = 1)	cisun				
Method	Sobol	EFAST	Morris	LHS SRRC	Random SRRC
Sensitivity decreases from top to bottom	tl	tl	tl	tl	tl
	btran	btran	ei	co2	co2
	daylfactor	daylfactor	btran	btran	btran
	ei	ei	co2	daylfactor	daylfactor
	co2	co2	daylfactor	ei	ei
	ea	ea	ea	ea	ea

Output (PFT = 1)	vcmxsun				
Method	Sobol	EFAST	Morris	LHS SRRC	Random SRRC
Sensitivity decreases from top to bottom	tl	tl	tl	tl	tl
	btran	btran	daylfactor	btran	btran
	daylfactor	daylfactor	btran	daylfactor	daylfactor

Table 5. Rankings for Needleleaf Evergreen Tree - Temperate

Chapter 9

Results: Broadleaf Evergreen Tree - Temperate (PFT = 5)

The same five methods of sensitivity analysis were tested on data pertaining to needleleaf evergreen trees in temperate zones. The 5 possible outputs were psnsun, rssun, cisun, lncsun, and vcmxsun. Again, the lncsun output was constant regardless of input and sensitivity analysis could not be performed on the lncsun output. Therefore, the lncsun data is omitted.

The bar graphs of first order and total order sensitivity indices from the Sobol Method on psnsun, rssun, cisun, and vcmxsun are the following:

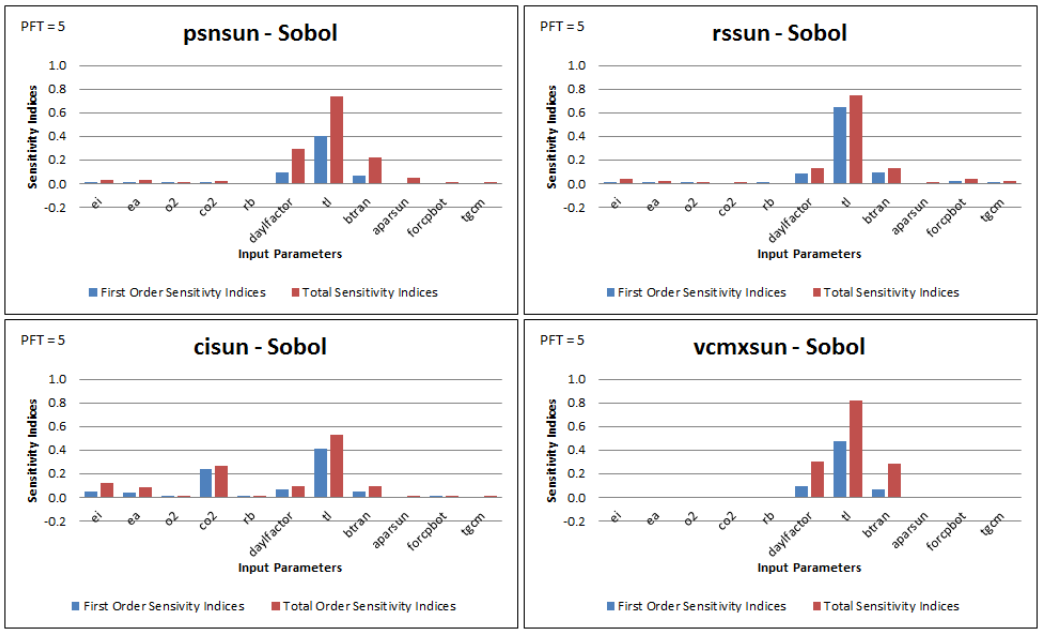


Figure 19. The Sobol Method

The bar graphs of first order and total order sensitivity indices from Extended FAST Method on psnsun, rssun, cisun, and vcmxsun are the following:

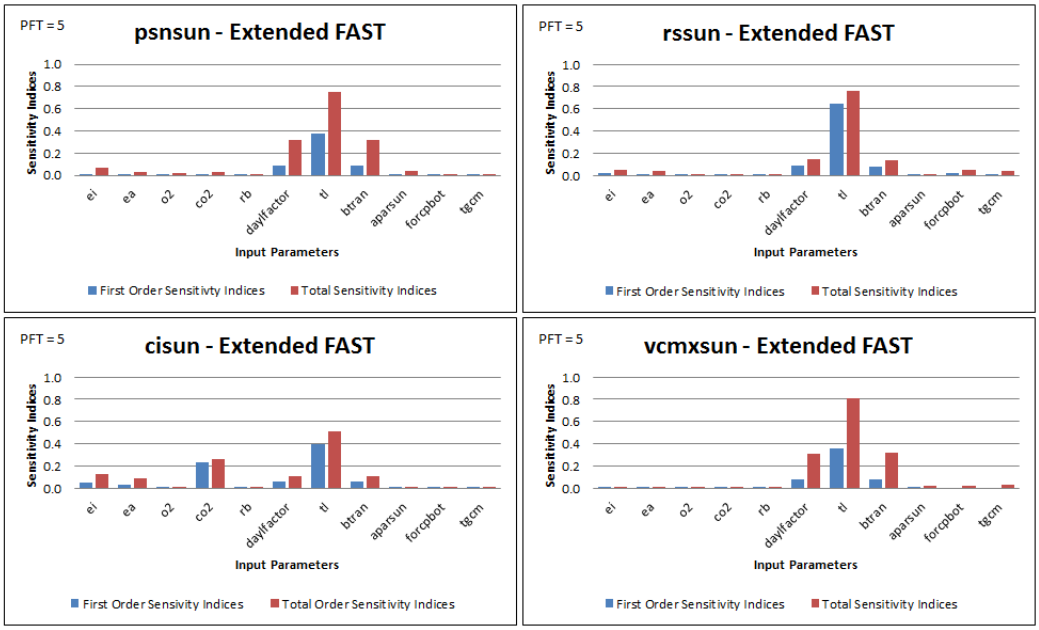


Figure 20. The Extended FAST Method

The bar graphs of sensitivity measures (mean and standard deviations of elementary effects) from the Morris Method on psnsun, rssun, cisun, and vcmxsun are the following:

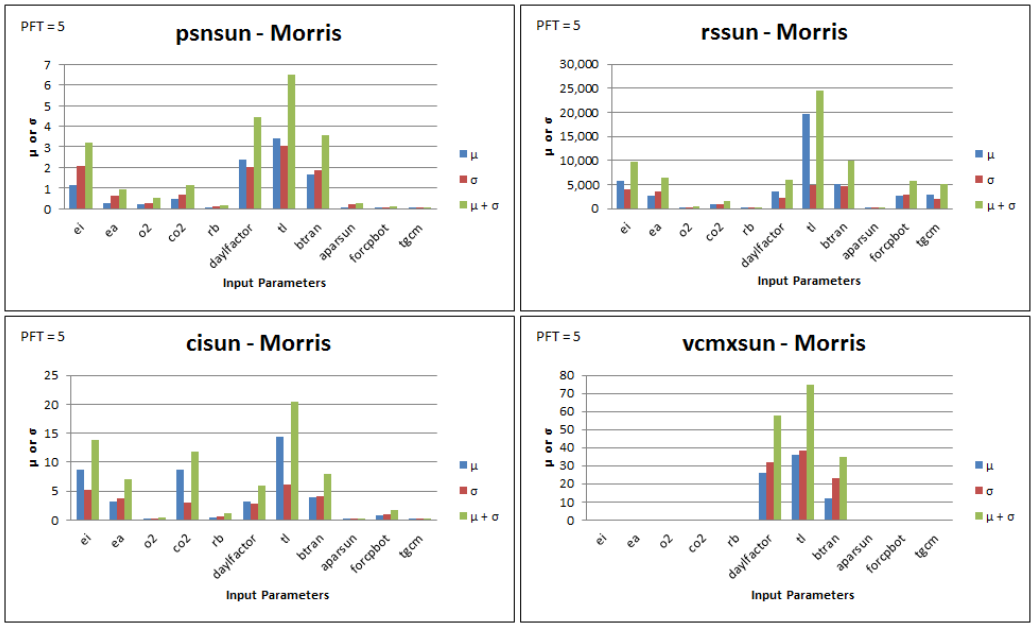


Figure 21. The Morris Method

The bar graph of sensitivity measures (SRRC) from Regression Analysis with LHS Sampling on psnsun, rssun, cisun, and vcmxsun is the following:

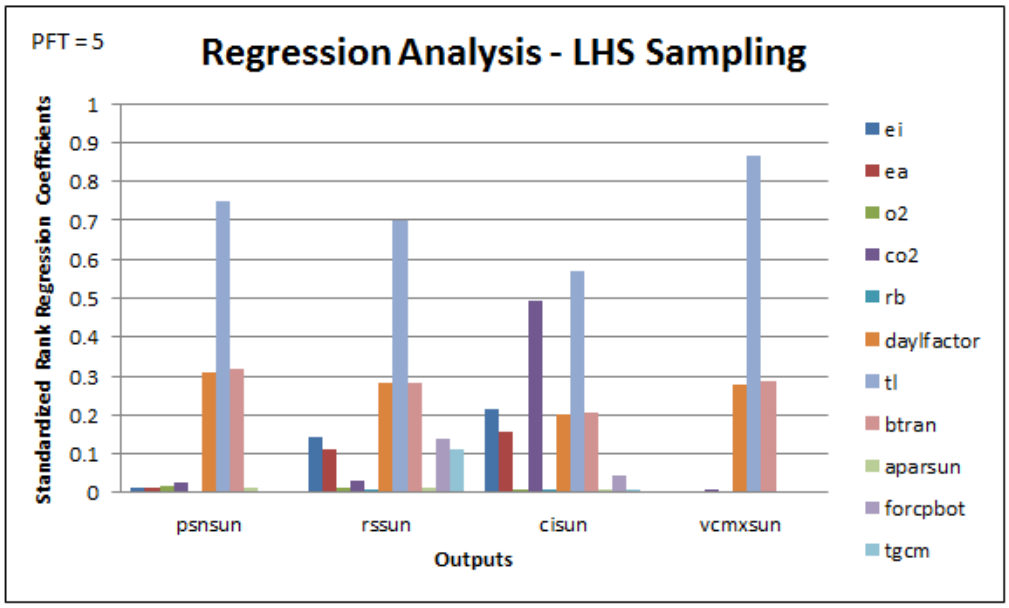


Figure 22. Regression Analysis based on Latin Hypercube Sampling

The bar graph of sensitivity measures (SRRC) from Regression Analysis with LHS Sampling on psnsun, rssun, cisun, and vcmxsun is the following:

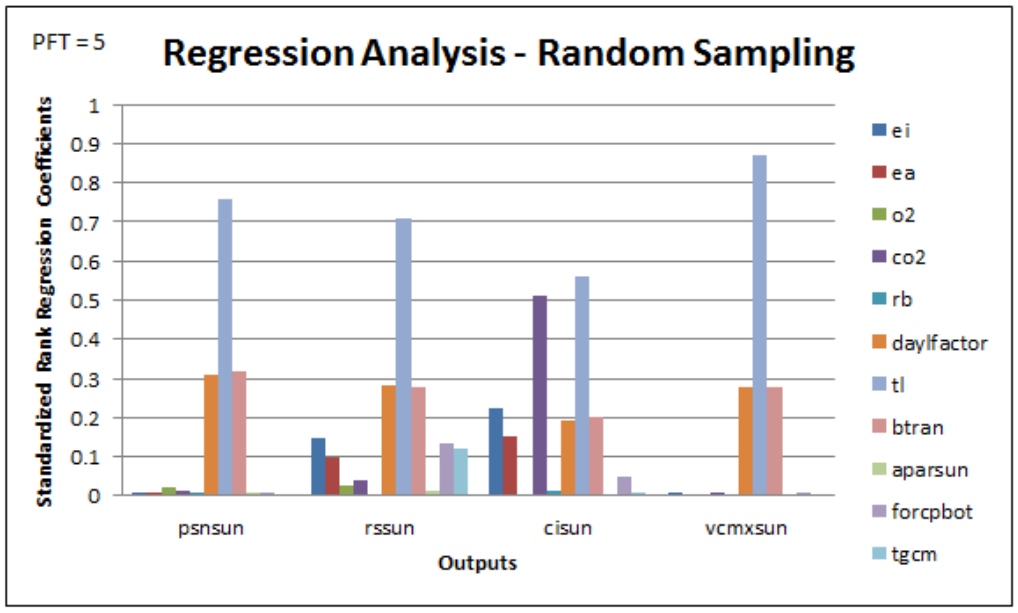


Figure 23. Regression Analysis based on Random Sampling

After producing the graphs and comparing them, a summary of the rankings based on the five methods were compiled in the tables below:

Output (PFT = 5)	psnsun				
Method	Sobol	EFAST	Morris	LHS SRRC	Random SRRC
Sensitivity decreases from top to bottom	tl	tl	tl	tl	tl
	daylfactor	btran	daylfactor	btran	btran
	btran	daylfactor	btran	daylfactor	daylfactor
			ei		
			co2		
			ea		

Output (PFT = 5)	rssun				
Method	Sobol	EFAST	Morris	LHS SRRC	Random SRRC
Sensitivity decreases from top to bottom	tl	tl	tl	tl	tl
	daylfactor	daylfactor	btran	btran	daylfactor
	btran	btran	ei	daylfactor	btran
			ea	ei	ei
			daylfactor	forcpbot	forcpbot
			forcpbot	tgcm	tgcm
			tgcm	ea	ea

Output (PFT = 5)	cisun				
Method	Sobol	EFAST	Morris	LHS SRRC	Random SRRC
Sensitivity decreases from top to bottom	tl	tl	tl	tl	tl
	co2	co2	ei	co2	co2
	ei	ei	co2	ei	ei
	btran	btran	btran	btran	btran
		daylfactor	ea	daylfactor	daylfactor
		ea	daylfactor	ea	ea
			forcpbot		

Output (PFT = 5)	vcmxsun				
Method	Sobol	EFAST	Morris	LHS SRRC	Random SRRC
Sensitivity decreases from top to bottom	tl	tl	tl	tl	tl
	daylfactor	btran	daylfactor	btran	btran
	btran	daylfactor	btran	daylfactor	daylfactor

Table 6. Broadleaf Evergreen Tree - Temperate

Chapter 10

Conclusion

This work aimed to achieve a couple of goals. The main goal was to determine which parameters are important and influential by implementing sensitivity analysis on a leaf photosynthesis-stomatal resistance model. Other goals include learning background information about sensitivity analysis and the computational model itself, learning how to use SimLab, learning how to design sensitivity analysis experiments, and how to analyze the results. Because of the mathematical nature of sensitivity analysis, some of the frameworks from probability and statistics used in developing sensitivity analysis methods were explored.

Chapters 7, 8, and 9 presented many results from the sensitivity analysis in the forms of graphs and tables. By comparing the tables visually, rankings for parameter based on the overall synthesis of all 5 methods were developed. These rankings are presented in the following:

Overall Synthesis (PFT = 7)				
Output	psnsun	rssun	cisun	vcmxsun
Sensitivity decreases from top to bottom	tl	tl	tl	tl
	btran	daylfactor	co2	btran
	daylfactor	btran	ei	daylfactor
	ei	ei	btran	
		forcpbot	daylfactor	
		tgcm	ea	
		ea		

Overall Synthesis (PFT = 1)				
Output	psnsun	rssun	cisun	vcmxsun
Sensitivity decreases from top to bottom	tl	tl	tl	tl
	daylfactor	btran	btran	btran
	btran	daylfactor	daylfactor	daylfactor
	ei	ei	co2	
	rb	forcpbot	ei	
	ea	tgcm	ea	
		ea		

Overall Synthesis (PFT = 5)				
Output	psnsun	rssun	cisun	vcmxsun
Sensitivity decreases from top to bottom	tl	tl	tl	tl
	btran	daylfactor	co2	btran
	daylfactor	btran	ei	daylfactor
	ei	ei	btran	
		forcpbot	daylfactor	
		tgcm	ea	
		ea		

Table 7. Overall rankings across all methods for each plant functional type

The first important conclusion that can be made is that tl (vegetation temperature) is the most sensitive and thus most important parameter across all methods and outputs. By inspecting all the graphs from Figure 9 to Figures 23, one can see that tl had much higher sensitivities than all other parameters. For example, the total sensitivity indices for tl had more than twice the value of those for the second ranked input under the psnsun and vcmxsun outputs and about more than four times the value of the second ranked input under the rssun output. This trend can essentially be seen on the other

two plant functional types and the Extended FAST method graphs as well (Figures 9, 10, 14, 15, 19, and 20).

The tl parameter also is dominating in the Morris Method graphs and the Regression Analysis graphs. In particular, the standardized rank regression coefficients for tl are more than twice the values of those for the second ranked input under the psnsun, rssun, and vcmxsun outputs regardless of sampling method.

The next conclusion to be made is which parameter ranked second highest. The two most important parameters after tl were btran (soil water transpiration factor) and daylfactor (daylength) for psnsun, rssun, and vcmxsun as can be inferred from Table 7. The co2 (atmosphere carbon dioxide concentration) was second specifically for cisun. This can be seen in each of the graphs for all plant functional types. The rankings for btran and daylfactor were not significantly different but btran usually was more sensitive than daylfactor by a very small amount. Since the psnsun output refers to the leaf photosynthesis rate and rssun refers to the stomatal resistance, then it can be regarded that btran and daylfactor both have essentially the same level of importance. The co2 having a more important role for cisun is in agreement with the fact that cisun signifies the sunlit intracellular carbon dioxide concentrations.

After tl, btran, and daylfactor, the next most important factor was ei (vapor pressure inside the leaf) which has a more prominent sensitivity measure compared to the lesser ranked parameters for the Morris Method as shown in Figure 11. The parameter ei is generally about half significant compared to btran and daylfactor for the Sobol and Extended FAST method as shown in Figure 15.

After ei, the lesser important parameters were generally forcpbot, tgcm, and ea in different orders depending on the output and plant functional type. The least important parameters can thus be inferred to be aparsun, o2, co2, rb, and to some degree forcpbot and tgcm. Specifically, the two least important parameters across all methods, plant functional types, and outputs are aparsun

(photosynthetically active radiation absorbed per leaf area index) and o2 (atmospheric oxygen concentration).

Besides the ranking of parameters, one can make the observation that the sensitivity analysis methods agree with each other for the most part because many conclusions about the parameters hold true across all sensitivity analysis methods. This can be supported by the fact that tl ranked first, btran and daylfactor and had very similar sensitivities, and co2 was ranked second under cisun across all sensitivity analysis methods generally.

As for the plant function types, the broadleaf deciduous tree in temperate zones (PFT = 7) and the broadleaf evergreen tree in temperate zones (PFT = 5) both had identical overall rankings as shown in Table 7. However, the needleleaf evergreen tree in temperate zones (PFT = 1) had a similar ranking as well. This ranking differed from the other two in regards to btran and daylfactor ranking differences, co2 ranking lower, and rb ranking higher.

There are still some more improvements that can be made in this work. For example, another step in the work would be to explore more in depth the biological basis of the rankings. Knowing that tl ranked first, more study can be undertaken to see how vegetation temperature plays a role in leaf photosynthesis rate and stomatal resistance. More biological background in the other parameters would be sought. This kind of study would also apply to the plant functional types. One would need to research into how these photosynthetic processes perform in different leaves and climate zones since it was noted that the broadleaf trees had identical rankings while the needleleaf tree did not. Moreover, there were other plant functional types with some involving different zones such as tropical or arctic zones that were not tested. These plant functional types should be tested as well.

Another type of improvement to this work would be standardizing the approach to compare all the sensitivities across output, plant functional types, and sensitivity analysis methods. This could involve setting an appropriate threshold. A more rigorous means of comparison should also be

investigated. Additionally, the probability distributions were assumed to be uniform for the parameter because not much was known about the parameters. There should be more inquiry in determining appropriate distributions for the parameters in future improvements.

From this work, one can see that sensitivity analysis proved to be a useful method in achieving the goal of determining which parameters are most important and sensitive. This work involved integrating biology, mathematics, statistics, and engineering and applying sensitivity analysis to a small model of leaf photosynthesis and stomatal conductance that is part of larger model that simulates the Earth's climate. On a larger scale, scientists, mathematicians, and engineers from all disciplines are working together to solve the problem of global warming. For example, the research team from [26] which provided the leaf photosynthesis and stomatal conductance subroutine used in this work is involved in improving the Community Land Model through linking the model with site measurements done by field experimentalists and exploring how to use sensitivity analysis to for model improvement. After conducting sensitivity analysis, one can begin to make choices concerning for instance how much resources should be allocated in studying certain parameters or even deciding which parameters can be excluded from the model for simplification. Therefore, sensitivity analysis can be a powerful tool for decision making especially at the public policy level involving cases such as global warming.

References

1. Box, G. E., & Draper, N. R. (1987*). Empirical model-building and response surfaces*. Oxford: John Wiley & Sons.
2. Smith, E. D., Szidarovszky, F., Karnavas, W. J., & Bahill, A. (2008). Sensitivity analysis, a powerful system validation technique. *Open Cybernetics & Systemics Journal, 2*, 39-56.
3. University Corporation for Atmospheric Research. (2013). *ABOUT CESM*. Retrieved from http://www2.cesm.ucar.edu/about
4. University Corporation for Atmospheric Research. (2013). *COMMUNITY LAND MODEL*. Retrieved from http://www.cesm.ucar.edu/models/clm/
5. University Corporation for Atmospheric Research. (2013). *CLM: SURFACE HETEROGENEITY*. Retrieved from http://www.cesm.ucar.edu/models/clm/surface.heterogeneity.html
6. Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., ... & Yang, Z. L. (2003). The common land model. *Bulletin of the American Meteorological Society*, *84*(8).
7. Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., ... & Slater, A. G. (2011). Parameterization improvements and functional and structural advances in version 4 of the Community Land Model. *Journal of Advances in Modeling Earth Systems*, *3*(1).
8. Wilson, T.V. (2006). *How the Earth Works*. Retrieved from http://science.howstuffworks.com/environmental/earth/geophysics/earth3.htm
9. University of California Museum of Paleontology. (2014). Retrieved from http://evolution.berkeley.edu/evolibrary/article/mcelwain_02
10. Farquhar, G. D., von Caemmerer, S. V., & Berry, J. A. (1980). A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species. *Planta*, *149*(1), 78-90.
11. Collatz, G. J., Ribas-Carbo, M., & Berry, J. A. (1992). Coupled photosynthesis-stomatal conductance model for leaves of C4 plants. *Functional Plant Biology*, *19*(5), 519-538.

12. Ball, J.T. (1988). *An analysis of stomatal conductance*. Ph.D. thesis, Stanford University.
13. Sellers, P. J., Randall, D. A., Collatz, G. J., Berry, J. A., Field, C. B., Dazlich, D. A., ... & Bounoua, L. (1996). A revised land surface parameterization (SiB2) for atmospheric GCMs. Part I: Model formulation. *Journal of climate*, *9*(4), 676-705.
14. Oleson, K. W., Lawrence, D. M., Gordon, B., Flanner, M. G., Kluzek, E., Peter, J., ... & Zeng, X. (2010). Technical description of version 4.0 of the Community Land Model (CLM). *NCAR Technical Note NCAR/TN-478+STR*.
15. Marino, S., Hogue, I. B., Ray, C. J., & Kirschner, D. E. (2008). A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of theoretical biology*, *254*(1), 178-196.
16. Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice: a guide to assessing scientific models*. John Wiley & Sons.
17. Saltelli, A. (1999). *Scheme for sensitivity analysis*. [Image]. Retrieved from http://upload.wikimedia.org/wikipedia/en/0/01/Sensitivity_scheme.jpg
18. Wainwright, H. M., Finsterle, S., Jung, Y., Zhou, Q., & Birkholzer, J. T. (2013). Making sense of global sensitivity analyses. *Computers & Geosciences*.
19. Confalonieri, R., Bellocchi, G., Bregaglio, S., Donatelli, M., & Acutis, M. (2010). Comparison of sensitivity analysis techniques: a case study with the rice model WARM. *Ecological Modelling*, *221*(16), 1897-1906.
20. Stewart, J. (2005). *Calculus: Concepts and Contexts* (3rd ed.). Belmont, CA: Brooks/Cole.
21. Ghahramani, S. (2005). *Fundamentals of Probability with Stochastic Processes* (3rd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
22. Jacques, J., Lavergne, C., & Devictor, N. (2006). Sensitivity analysis in presence of model uncertainty and correlated inputs. *Reliability Engineering & System Safety*, *91*(10), 1126-1134.

23. Confalonieri, R., Bellocchi, G., Bregaglio, S., Donatelli, M., & Acutis, M. (2010). Comparison of sensitivity analysis techniques: a case study with the rice model WARM. *Ecological Modelling*, *221*(16), 1897-1906.
24. Giglioli, N., & Saltelli, A. (2000). SimLab 1.1, Software for Sensitivity and Uncertainty Analysis, tool for sound modelling. *arXiv preprint cs/0011031*.
25. Tarantola, S. (2005). SimLab 2.2 Reference Manual. *Institute for Systems, Informatics and Safety, European Commission, Joint Research Center, Ispra, Italy*.
26. Wang, D., Xu, Y., Thornton, P., King, A., Steed, C., Gu, L., & Schuchart, J. (2014). A functional test platform for the Community Land Model. *Environmental Modelling & Software*, *55*, 25-31.
27. Joint Research Centre of the European Commission (2014). *Software SIMLAB*. Retrieved from http://ipsc.jrc.ec.europa.eu/?id=756
28. Yang, J. (2011). Convergence and uncertainty analyses in Monte-Carlo based sensitivity analysis. *Environmental Modelling & Software*, *26*(4), 444-457.
29. Sobol', I. Y. M. (1990). On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe Modelirovanie*, *2*(1), 112-118.
30. Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, A. G., & Schaibly, J. H. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. *The Journal of Chemical Physics*, *59*(8), 3873-3878.