AN ABSTRACT OF THE DISSERTATION OF

Marcela Brugnach for the degree of Doctor of Philosophy in Bioresource Engineering and Forest Science presented on December 19, 2002.

Title: New Tools and Approaches to Uncertainty Estimation in Complex Ecological Models

Abstract approved

Redacted for privacy

John Bolte

Redacted for privacy

Gay Bradshaw

This dissertation investigates the problem of uncertainty in complex ecological models. The term "complex" is used to convey both the common and scientific meanings. Increasingly, ecological models have become complex because they are more complicated; ecological models are generally multi-variate and multi-leveled in structure. Many ecological models are complex because they simulate the dynamics of complex systems. As a result, and as science moves from the modern/normal to postmodern/post-normal paradigm view of the world, the definition of uncertainty and the problem of uncertainty estimation in models tread the lines between the technical and the philosophical. With this in mind, I have chosen to examine uncertainty from several perspectives and under the premise that the needs and goals of uncertainty estimation, like ecological models themselves, are evolving. Each chapter represents a specific treatment of uncertainty and introduces new methodologies to evaluate the nature, source, and significance of model uncertainty. In the second chapter, '*Determining the significance of threshold values uncertainty in rule-based classification models*', I present a sensitivity analysis methodology to determine the

significance of uncertainty in spatially-explicit rule-based classification models. In the third chapter, '*Process level sensitivity analysis for complex ecological models*', I present a sensitivity analysis methodology at the process level, to determine the sensitivity of a model to variations in the processes it describes. In the fourth chapter, '*A Component Based Approach for the Development of Ecological Simulations* ', I investigate how the process of developing an ecological simulation can be advanced by using component-based simulation frameworks. I conclude with reflection on the future of modeling and studies of uncertainty.

New Tools and Approaches to Uncertainty Estimation in Complex Ecological Models

by

Marcela Brugnach

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of

the requirements for the

degree of

Doctor of Philosophy

Presented December 19, 2002

Commencement June 2003

Doctor of Philosophy dissertation of Marcela Brugnach presented on December 19, 2002

APPROVED:

Redacted for privacy

Major Professor, representing Bioresource Engineering

Redacted for privacy

Chair of the Department of Bioengineering

Redacted for privacy

Major Professor, representing Forest Science

Redacted for privacy

Chair of the Department of Forest Science

Redacted for privacy

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Redacted for privacy

Marcela Brugnach, Author

ACKNOWLEDGEMENTS

I would like to thank my advisors Gay Bradshaw and John Bolte, as well as the members of my committee Ajoy Velayudhan, Kermit Cromack, James Reeb, and the recent participation of Steve Radosevich and Marshal English.

I also would like to thank:

Ronald P. Neilson and the people from his lab, Dominique Bachelet, Lisa Baldhuman, Jim Lenihan and Ray Drapek.

My friends and colleagues Jill Anthony, Giovanna and Jesus Aleman, Charles Hillyer, Cyndy Hines, Hamidou Sy and the late Assinatou Johnson, Marcelo and Roberta Fernandez, Kathryn Ronnenberg, María Fiorella, Mercedes Berterretche, Jeff Borchers, Claudia and Sebastián Pierini, Sarah Hoffman, Scott Waichler, Pat Bergen, France Lamy, Monica Zamit, Sandra and Daniel Ortíz, Nan Johnson.

Susan Ustin and the people from the CSTARS lab at UC Davis.

Finally, a big thank to my family, for their understanding and support. To my bright husband, Pablo, and my beautiful daughter, Manuela. To my parents, Dora and Antonio, my sister Daniela and my niece Jazmín.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF TABLES

LIST OF TABLES (continued)

New Tools and Approaches to Uncertainty Estimation in Complex Ecological Models

Chapter 1

Introduction

Uncertainty: Normal Science and Post-Normal Science Views

If there is one element which characterizes today's environmental problems from past decades, it is its apparent complexity. Science has always been preoccupied with the c*omplicated* essence of nature; the underlying belief was that with more data, faster computers, and more detailed models we would be able to solve the complicated problems and questions that the environment posed. However, now, with technological capacities and volumes of information, nothing much has gotten easier in terms of solving ecological questions of description and prediction. Against the optimistic predictions of traditional science, the uncertainty associated with our attempts to explain and solve environmental problems does not seem to have been substantially reduced.

This situation has forced scientists to re-evaluate the magnitude of the task of understanding natural systems (Funtowicz and Ravetz, 1990; Funtowicz and Ravetz, 1993). As these systems were more critically observed, the perception of its intrinsic complexity (as opposed to merely complicatedness) arose. Natural systems are composed by subsystems, which are combined and interact as a result of an organization that is intrinsic to the system, thus called "self-organization" (Cilliers,1998). This property, combined with the ability to "read" or interpret the information from their surrounding (often called "representation") allows them to adapt themselves to changes through acclimation or learning. The organization of the subsystems is often hierarchical and the interaction between components, non-linear. These properties make the understanding and prediction of systems particularly difficult.

When dealing with environmental problems, an additional factor makes the study of the systems involved particularly complex. This has to do with the realization that human beings with their beliefs and perceptions about nature are part of the systems to be studied. So, even the identification and conceptualization of a particular aspect of reality to be solved, may influence the type of prediction obtained (Funtowicz and Ravetz, 1990; Funtowicz and Ravetz, 1993; Allen et al., 2000).

The uncovering of complexity in nature also made scientists scrutinize many current assumptions about the way traditional (also called "Normal", Funtowicz and Ravetz, 1993) science is done. Traditional science relies on analytical methods to understand these problems. In Normal science's protocol, intrinsic complexity of problems is reduced to elements that can be studied in isolation in more detail. Traditional science often tends to be reductionist. Normal science practitioners favor theories that have simpler formulations and that are based on fundamental, permanent principles, usually closer to the realm of physics or chemistry (Chargaff, 1997). Causal relationships between the elements are determined, and the basic principles are extracted. Elements of the problem can then be reassembled into the building of a more comprehensive, general theory. This process of theory construction is thought to be cumulative and thus, if done properly, it gradually works towards increasing understanding and reducing uncertainties about the problem. The implicit idea behind cumulative knowledge is that objects of study are part of a reality for which there is an explanation (or a theory) that properly describes it. Thus science seeks to come progressively closer to the truth by finding the single, best explanation possible: one that is more and more free of uncertainties.

Complex systems, however, operate under conditions far from equilibrium. Their behavior frequently may appear not driven by simple linear causalities, and their alternative developmental paths may drastically diverge (Weaver, 1948; Cilliers, 1998). These properties make complex systems intrinsically uncertain, relative to the levels of certainty sought by modern science, and they often can be described in more than one possible way. Therefore, the traditional approaches of Normal science: reductionism and analysis, together with the assumption that there is always one theory that is closest to the truth (cumulative knowledge), may not be the most adequate way of confronting complex systems (Allen et al., 2000).

Even when it may appear that some understanding of a complex problem has been achieved, it may be difficult to determine whether the concomitant reduction in uncertainty is due to an actual understanding of the object of study or due to an oversimplification of the problem (Allen et al., 2000). When trying to understand complex problems, the urge to arrive to a single, comprehensive and universal explanation may force to neglect some aspects of the problem that do not fit in the general theory. In other words, in the urge to reduce uncertainty, the scope and its understanding is narrowed.

An alternative view to traditional science, often called post-Normal, post-structuralist or post-modern science has recently arisen. This view appears as a direct response to the apparent shortcomings of analytical and reductionist approaches when dealing with complex systems. It offers a more relaxed view of reality by stressing less on the necessity of an abstract meta-narrative that needs to be continuously validated by scientific knowledge. Post-modern science recognizes the validity of explanations,

not because they necessarily reflect a universal truth, but because they are consistent with a given conceptual context and successfully help to understand systems or phenomena. Post-modern science stresses on the importance of considering the intrinsic constraints of our knowledge to judge the reliability of our predictions. By doing so, it admits the possibility of more than a single, unique explanation or "narrative", and thus, it is more open to the incorporation of uncertainty within the framework of theories.

As we move to post modern times, the value of uncertainty is recognized, shifting from its elimination to a better understanding of its significance. From this perspective, uncertainty carries information that is valuable to point out which are the limits of our knowledge (Bradshaw and Borchers, 2000). By doing this, we eliminate the misleading conception of looking for the singular best explanation to welcome a plurality of legitimate explanations useful to reach a better understanding of the system. In this context, a new role for scientist is to assure the scientific information conveys uncertainty as best as possible; this includes managing uncertainty as well as providing new methodologies and tools that make it possible.

Simulation models and uncertainty

From the perspective of post-normal science, ecological simulations are tools that play a pivotal role in the study of complex systems. They are used to investigate the properties and causal aspects of ecological phenomena for which experimentation is not a feasible alternative. A simulation is a mechanism for describing interactions

among subsystems in a system and generating dynamical phenomena (Rasmussen and Barret, 1995). The dynamics of the system are not explicitly encoded in the simulation rather they emerge as a result of the collective interaction of subsystems. To construct an ecological simulation it involves identifying which subsystems are needed to characterize the ecosystem to be simulated; formulating a model for each subsystem and defining the interactions between these subsystems. Finally, a simulation is implemented in a computer representation to execute and evaluate it. Many different approaches can be used to build the models that participate in a simulation, ranging from statistical formulations to rule-based classification or highly parameterized process-based ones (Haefner, 1996). Each approach imprints different characteristics on the model structure, results, and the sources of uncertainty that affect it. Commonly used in ecological studies, and subject of study in this dissertation, are rule-based classification model and process-based models. Rule-based classification models represent ecological theory using rules and numerical thresholds to classify information. The numerical thresholds are used to define the boundaries of the classes while the rules are used as the inference mechanism. Process-based models simulate ecosystems by mimicking the functions of its component processes.

Although in principle building a simulation is a simple task, in practice, the nature of the problems for which many simulations are created to address, and the nature and limitations of the information available to solve the problem constitute significant sources of uncertainty for the development. Uncertainty is as complex as the model itself and cannot be seen as external to either the modeling process nor its analysis and evaluation. As defined by Zimmermann (2000) "uncertainty implies that in certain

situation a person does not dispose about information which quantitatively and qualitatively is appropriate to describe, prescribe or predict deterministically and numerical a system its behavior or other characteristic". This uncertainty provides information about the reliability of model predictions and limits our confidence in this predictions (Ravetz, 1999a; 1999b); therefore playing a pivotal role in model evaluation and interpretation. A good modeling practice requires understanding of the effects of uncertainty in model behavior.

Sources of uncertainty in simulation models

The construction of a simulation model embodies a series of nested, and sometimes iterative, assumptions and subjective decisions about the behavior of the ecosystem and its most relevant features (Zimmermann, 2000; Klir, 1999). All these issues incorporate uncertainty into the simulation model at some level.

Uncertainties have different characteristics and origins. One source of uncertainty arises from the intricacy of the information used (Zimmermann, 2000). Ecosystems are complex system and their description requires dealing with complex information. To deal with this issue, it is usually necessary to simplify the system description by focusing on those elements that are considered most important (Zeigler, 1984; Klir, 1991). Model construction, then, implicitly involves the embedding of assumptions at many levels, and the particular, subjective (although perhaps consensus mediated) beliefs of the modeler. These assumptions include system boundary delineation, the choice of temporal and spatial scales of description, and the selection of an appropriate level of detail.

Another source of uncertainty that arises in the model construction process relates to the lack of information, and sometimes lack of consensus, about the characteristics of the ecological phenomena to be simulated (Zimmermann, 2000). To deal with these types of situations, it is usually necessary to make further assumptions about how the system behaves and also to test the incorporation of several different alternatives in the model. Other sources of uncertainty in ecological modeling originate from measurements errors of physical properties and the natural spatial and temporal variability that characterizes the ecosystems (Katz, 1999).

Uncertainties are expressed in the different aspects of the model such as parameter values used, the structure of the model (i.e., the form of a model as defined by its components and their relationship's functional form), the level of detail incorporated, and the spatio-temporal scale at which the phenomena is represented.

Methods to assess uncertainty in simulation models

Numerous methods are currently used and available for assessing model uncertainty in ecological studies (see: Katz, 1999; Omlin and Reichert, 1999; Campolongo et al., 2000; Kann and Weyant, 2000 for general review). These methods are generally classified either as sensitivity or uncertainty analysis. In both types the central goal is to diagnose the effects that variations in input factors have on model behavior (Saltelli, 2000).

Sensitivity analysis is a general approach to study model behavior (Haefner, 1996; Elston, 1992). As defined by Saltelli et al. (2000) sensitivity analysis is an activity that

studies the relationship between information flowing in and out of the model. It aims to determine the rate of change in model output in response to variations in input factors (i.e., parameters or input data are varied; Katz, 1999). This variability typically represents the uncertainty associated with model parameters, structure, and assumptions, or the data fed to a model. The measure of model's sensitivity indicates how much an output can vary relatively to the variations in these input factors. High levels of sensitivity indicate that a small change in the input data or parameters values can generate a relatively large change in model results. Conversely, low sensitivity levels indicate that even large changes in input data or parameter values do not significantly affect model outcome. Some methods of sensitivity analysis are based on computing a measure of sensitivity (e.g., correlation analysis, linear regression, measures of importance, sensitivity indexes) where the effect changing in input factors have on model output is studied. However, rather than measuring uncertainty, these methods place their emphasis in apportioning variations of the output to the different input factors.

Uncertainty analyses measure the uncertainty of model's results. This class of analyses is concerned with estimating the overall uncertainty of model output given the uncertainty associated with parameters or input data (Saltelli et al., 2000). In other words, the uncertainty associated with the output represents what it is unknown about the ecosystem modeled. For example, a commonly used method, Monte Carlo uncertainty analysis (Saltelli et al., 2000), is based on a random sampling of the entire input factor space (e.g., parameters, input data) and determines how uncertainty propagates through the model and affects model output.

From the number of methods available, there is no a single one that can be considered general enough to be applicable to deal with all sort of uncertainties in ecological models. Each method has different uses and applications (Saltelli et al., 2000) and it conveys uncertainty information differently. The selection of method to use depends on the goals of the analysis, the characteristic of the model to analyze (e.g., modeling approach used to develop the model), and the aspects of uncertainty that need to be uncovered.

Research Objectives

Despite the availability of methods, all of them present some limitations to assess uncertainty in complex ecological models. One limitation is that these methods frequently overlook the fact that uncertainty varies between different ecosystems and spatial locations. This limitation becomes apparent when these methods are applied to spatial explicit classification models, where the output classification varies through space and represent diverse ecosystems. In this type of models, the interpretation and evaluation of the effects of uncertainty needs to be effected in a context where spatial and ecological information is considered simultaneously.

On the other hand, the characteristics that the modeling approach imparts to the phenomena represented cannot be ignored when determining the effect of uncertainty. In this regard, a rule-based approach to modeling also presents some shortcomings when applied to ecological problems. One of the main drawbacks is that the concepts represented by ecological models may be vague and imprecisely defined, however, in a rule-based model, each class in the model is meant to

represent a precisely defined theoretical concept. This problem generally is addressed by artificially delineating boundaries among the classes, thereby generating discretely defined classes with "crisp" boundaries. A consequence of this solution is that the classification in the vicinity of these artificially delineated boundaries (i.e., defined by numerical thresholds) may lead to invalid results. In these regions, small alteration of the numerical thresholds can generate a different classification output. Under these circumstances, modeling outcome cannot be interpreted as a precise description of ecological theory but rather as an outcome of the particular model representation used.

Chapter 2, '*Determining the significance of threshold values uncertainty in rule-based classification models',* has the objective of improving the interpretation and evaluation of spatially-explicit rule-based classification models. This paper presents a stepwise methodology that, using a genetic algorithm, determines the effects of uncertainty in model output and identifies the most likely alternative results. These alternative results are included as an integral part of the output. This methodology considers that the effects of uncertainty are relative to their spatial location and ecosystem under scrutiny and, based on that information it computes a measure of confidence in model results. This measure helps the modeler interpret and evaluate the significance uncertainty has in the model output. In this chapter, I present a case study applying the methodology to the global vegetation model Mapped Atmosphere-Plant-Soil System (MAPSS, Neilson 1995).

Another limitation of the uncertainty assessment methods is that they can be difficult to apply to process-based models characterized by a large parameter space. This is

because a large parameter space requires a large number of model runs to perform the analysis and the cost of the analysis can become too high. On the other hand, difficulties may arise when the analysis does not parallel the way in which a model is thought; that is, in terms of processes that interact among each other to generate a realistic representation of how the ecosystem works. This is important because understanding how the output is generated constitutes an essential part of the corroboration of a model and the assurance of its quality (Beck, 2002). Also, ignoring how a model is conceptualized and organized in terms of processes hinders the investigation of the effect process uncertainty may have in model behavior. Generally, process-level uncertainty is "black-boxed" and evaluation of model uncertainty is restricted to investigating model behavior by analyzing the relationships between input-output pair patterns. Thus, uncertainty intrinsic to the conceptualization of ecosystem processes is frequently avoided. Any variation derived from the process level is subsumed and confounded in the input-output analysis.

To overcome these difficulties, in Chapter 3, '*Process level sensitivity analysis for complex ecological models*', I propose a methodology of sensitivity analysis applicable to complex process-based models. This method focuses on sensitivity analysis at the process level; aiming to determine how sensitive the model output is and to what extent this sensitivity is due to variations in the processes described by the model. This approach differs from most traditional sensitivity analyses in which the focus is in apportioning output variability to input factors such as parameters or input data. The information by the analysis proposed here, can be used to ensure that the way in which the model operates resembles the phenomena being

modeled, when uncertainties in processes are considered. It allows evaluation of the quality of a model and increases the confidence in model results. Some of the advantages of this approach are that it allows handling high levels of complexity by abstracting information at the process level, it facilitates the interpretation of model behavior and it provides information that allows exploration of how uncertainty in processes might affect model output. I illustrate the method through an example using the vegetation model Biome-BGC (Thornton 1998, www.forestry.umt.edu/ntsg).

A third limitation that hinders the investigation of model uncertainty is the one that originates from *how* the simulation is implemented. Unfortunately, current modeling paradigms (e.g., procedural programming, object oriented programming) fail in offering good solutions for the understanding of uncertainty; often limiting *what* can be done to assess uncertainty. To this end in Chapter 4, '*A Component Based Approach for the Development of Ecological Simulations*', I investigate how component-based framework technology can advance this process by providing better tools to deal with the characteristics of ecological simulations, and thus facilitating the investigation of the many sources of uncertainty present in a simulation model. A component based simulation framework is a generic software structure that implements generalized behavior common to all simulations, which can be extended to create a more specific subsystem or application. It fosters component based software development as a programming paradigm. According to this paradigm a simulation is built as a collection of components, where each component is an executable entity that can be considered a standalone service

provider. In this paper I examine the ability of component based simulation frameworks to improve the modeling process in these areas:

1. effecting an incremental development of a new simulation and
2. modifying existing formulations so different modeling assumptions can be tested. These questions are explored through a case study using the simulation framework ModCom (Hillyer et al., 2002).

References

Allen, T.F.H., J.A. Tainter, J.C. Pirex and T.W. Hoekstra. 2000. Dragnet ecology- "Just the facts, Ma'am: The privilege of science in a postmodern world. BioScience. 51(6):475-485.

Beck, B. 2002. Model evaluation and performance. Encyclopedia of environmetrics. In El-Shaarawi, A.H. and W.W. Piergorsch. John Wiley & Sons, Ltd, Chichester, 3:1275-1279.

Bradshaw, G.A. and J. G. Borchers. 2000. Uncertainty as information: narrowing the science-policy gap. Conservation Ecology, 4(1):7. [online] URL://www.consecol.org/vol4/iss1/art7

Campolongo F., A. Saltelli, T. Sorensen and S. Tarantola. 2000. Hitchhiker's Guide to Sensitivity Analysis. *In* Sensitivity Analysis, Edited by A. Saltelli, K. Chan and E.M.Scott. Willey Series in probability and statistics, pp 15-47.

Chargaff, E. 1997. In dispraise of reductionism. BioScience. 47(11): 795-797.

Cilliers, P. 1998. Complexity and postmodernism. Understanding complex systems. Routledge, 154 pp.

Elston D. A. 1992. Sensitivity analysis in the presence of correlated parameter estimates. Ecological Modelling. 64:11-22.

Funtowicz, S. O. and J. R. Ravetz. 1990. Uncertainty and quality in science for policy. Kluwer Academic, Dordrecht, The Netherlands, 232 pp.

Funtowicz, S. O. and J. R. Ravetz. 1993. Science for the post-normal age. Futures. 25(7): 739-755.

Haefner, J. W. 1996. Modeling Biological Systems, principles and applications. Chapman and Hall, 473 pp.

Hillyer, C., J. Bolte, F. van Evert and A. Lamaker. 2002. The ModCom Modular Simulation System. Journal of European Agronomy. In press.

Kann, A. and J.P. Weyant. 2000. Approaches for performing uncertainty analysis in large scale energy/economic policy models. Environmental Modeling and Assessment. 5(2000) 29-46.

Katz, R. W. 1999. Techniques for estimating uncertainty in Climate Change scenarios and Impact studies. In: Representing Uncertainty in Climate Change Scenarios and Impact Studies. Proceeding of the EU Concerted Action Initiative ECLAT-2 Workshop 1, Helsinki, Finland, April 16th - 18th, 1999, Climatic Research Unit, Norwich, UK., pp 38-53.

Klir, G.J. 1991. Facets of systems science. Plenum Press, New York, 664 pp.

Klir, G.J. and M.J. Wierman. 1999. Uncertainty-based information. Elements of generalized information theory. Physica-Verlag, 168 pp.

Neilson, R.P. 1995. A model for predicting continental scale vegetation distribution and water balance. Ecological Applications, 5:362-386.

Omlin, M. and P. Reichert. 1999. A comparison of techniques for the estimation of model uncertainty. Ecological Modelling, 115:45-59.

Rasmussen, S. and C.L.Barret. 1995. Elements of a Theory of Simulation. Santa Fe Institute, working paper 95-04-040.

Ravetz, J. 1999a. Models as Metaphors, ULYSSES Working Paper WP-99-3, Darmstadt, University of Technology.

Ravetz, J. 1999b. Developing principles of good practice in integrated environmental assessments. International Journal of Environment and Pollution.11(3): 243-265.

Saltelli, A. 2000. What is Sensitivity Analysis? *In* Sensitivity Analysis, Edited by A. Saltelli, K. Chan and E.M.Scott. Willey Series in probability and statistics, pp 3-13.

Thornton, P.E. 1998. Description of a numerical simulation model for predicting the dynamics of energy, water, carbon and nitrogen in a terrestrial ecosystem. Ph.D. dissertation. University of Montana, Missoula, MT 280 pp. [Available from Mansfield Library, University of Montana Missoula, MT 59812].

Weaver, W. 1948. Science and complexity. American Scientist. 36:536-544.

Zeigler, B.P. 1984. Theory of modelling and simulation. Krieger Publishing Company,Malabar, Florida, 435 pp.

Zimmermann, H.-J. 2000. An application oriented view of modeling uncertainty. European Journal of Operational Research, 122:190-198.

Chapter 2

Determining the significance of threshold values uncertainty in rule-based classification models

Ecological Modelling
P.O. Box 181
1000 AD Amsterdam
The Netherlands

This paper is currently in press

Introduction

Today, ecological models play an increasingly important role in scientific and decision-making arenas. As the spatio-temporal frames extend to national and global scales and socio-ecological issues increase in complexity, models are an integral tool for providing insight into understanding ecosystem dynamics and a way to develop viable environmental policy. Subsequently, ecological models represent a key source of information for policy-makers whose decisions must reflect the most current and accurate state of knowledge. However, the complexity of these models and their intrinsic uncertainties presents a challenge for accurately interpreting modeling results.

Uncertainty refers to what it is *unknown* about the ecosystem being modeled, which affects the appropriate description and prediction of the ecosystem behavior. A model is subject to many sources of uncertainty, including the error of measurements in parameters or input data, the lack of information about the ecosystem characteristics and the poor understanding of its underlying mechanisms (Saltelli, 2000). These sources of uncertainty not only restrict what can be modeled, but, by propagating through the model, they also impose a severe limitation in the model results. While the uncertainties associated with model output are generally ignored, they can provide substantive information (Morgan and Henrion, 1990; Klir and Wierman, 1999). Model output uncertainty provides information about the precision of the output, the range of possible alternatives, and confidence associated with a model's predictions. From this perspective, output uncertainty carries as much information as does the model output; uncertainty is a source of information (Klir and

Wierman, 1999; Bradshaw and Borchers, 2000). Assessing model output uncertainty shifts from the goal of its elimination to better understanding its significance and interpretation.

Numerous methods are currently used and available for assessing model uncertainty in ecological studies (see Katz, 1999; Omlin and Reichert, 1999; Campolongo et al., 2000; Kann and Weyant, 2000 for general review). These methods are generally classified either as sensitivity or uncertainty analysis where the central goal is to diagnose the effects of variations in parameters or input data on model behavior (Saltelli 2000). Each of these methods offers its own particular perspective and imparts a certain influence on inference. For example, commonly used is the Monte Carlo method of uncertainty analysis, which based on a random sampling of the entire input factor space (e.g., parameters, input data) determines how uncertainty propagates through the model and affects model output. In contrast, methods of sensitivity analysis based on computing a measure of sensitivity (e.g., correlation analysis, linear regression, measures of importance, sensitivity indexes, measure of importance) study the effect changing in input factors have on model output. Rather than measuring uncertainty, these methods place their emphasis in determining the extent of the contribution of input factors uncertainty to output uncertainty. All these sensitivity and uncertainty analysis methods, however, frequently fail in assessing uncertainty when its effects vary within different ecosystems. In addition, the spatial context or spatial dependency in the measurement of uncertainty is seldom contemplated. These characteristics present a problem when these methods are applied to spatial models, where interpretation and evaluation of the effects of

uncertainty can only be effected in a context where spatial and ecological information are considered simultaneously.

The sources of uncertainty depend on the way in which the given model has been conceptualized. Many different approaches have been used to conceptualize model hypotheses, ranging from statistical formulations to rule-based classification or highly parameterized process-based models (Haefner, 1996). Each approach imprints different characteristics on the model structure, results, and sources of uncertainty.

Commonly used in ecological studies, rule-based classification models represent ecological theory using rules and numerical thresholds to classify information. The numerical thresholds are used to define the boundaries of the classes while the rules are used as the inference mechanism. Each class in the model is meant to represent a precisely defined theoretical concept. The inputs of the model are the attributes that characterize the object(s) to be classified. In a spatially explicit model, each object also has a spatial location in a landscape. Once the model classifies all objects, the output is represented by a spatial configuration of classes. One of the main drawbacks of applying this modeling paradigm to ecological problems is that the concepts represented by ecological models may be vague and imprecisely defined. This problem generally is addressed by artificially delineating boundaries among the classes, thereby generating discretely defined classes with "crisp" boundaries. A consequence of this solution is that the classification in the vicinity of these artificially delineated boundaries (i.e., defined by numerical thresholds) may lead to invalid results. In these regions, small alteration of the numerical thresholds can generate a different classification output. Under these circumstances, modeling

results cannot be interpreted as a precise description of ecological theory but rather as an outcome of the particular model representation used.

We present a methodology to determine the significance of uncertainty in rule-based classification models. Our goal is to improve model interpretation and evaluation by explicitly incorporating the effects of uncertainty as part of the model output. Within this framework, uncertainty can be defined in terms of the degree to which assumptions about threshold values propagate through the model and affect its output. Specifically, we seek to determine what the distribution of potential outcome classes might be when uncertainty in threshold values is propagated through the model. In some cases, more than one possible classification is generated. Uncertainty is explicitly acknowledged by including the most likely classifications as an integral part of model output. Evaluation and interpretation of model results are carried out using a set of ecologically-based metrics that allow analysis of the output from an ecological perspective. In this article, the methodology is applied to the biogeographical model MAPSS (Neilson 1995), a model that assigns potential vegetation classes to spatial locations in response to landscape and climatic drivers.

Methodology

The methodology presented here involves three steps: (1) determination of the effects of threshold uncertainty in model output, (2) interpretation of model results, and (3) incorporation of alternative results into model output. This is applied to gridded data, where each grid cell represents a spatial location on a landscape.

Step 1: Determination of the Effects of Threshold Uncertainty in Model Output

To determine the effects of threshold uncertainty on model output, the methodology determines how the model responds to perturbation in the threshold values. It is assumed that the threshold values used in the model are the best fit estimates and that they provide an optimal representation of reality. In this context, uncertainty associated with these best estimates of threshold values is represented by variability associated with the estimate. This first step is carried out by perturbing the threshold values using their variability estimates to determine, via sensitivity analysis, the maximum sensitivity value for each spatial location in the landscape and by identifying the most likely alternative model results.

Maximum sensitivity value

The main goal here is to determine the maximum sensitivity value for each spatial location. Sensitivity is defined as the rate of change of the model output as input parameters (i.e., threshold values) are changed (Saltelli, 2000). High values of sensitivity indicate that a small change in threshold values can generate a relatively large change in model results. Conversely, low values of sensitivity indicate that even large changes in the threshold values do not significantly affect model outcomes. Changes in output are measured with the dissimilarity index presented by Sykes et al. (1999).

Consider, for example, a vegetation classification model. The dissimilarity index compares differences between output vegetation classes. Different applications may

require adapting the index to the specific model outcomes of interest. The dissimilarity index compares vegetation classes by considering the relative importance of different plant life forms in each vegetation class as well as attributes of each life form (i.e., deciduous, broadleaf, c3), weighting each attribute. A vegetation class represents a large community of plants composed of one or more life forms (i.e., tree, grass, shrub). The dissimilarity index is described by:

$$\Delta V(i,j) = 1 - \Sigma_k \{\min (V_{ik}, V_{jk}) * [1 - \Sigma_l w_{kl} |a_{ikl} - a_{jkl}|]\}$$

where,

$\Delta V(i,j)$ is the dissimilarity between vegetation classes i and j

V_{ik}, V_{jk} are the importance values of plant life form k in vegetation class i and j

$a_{ikl} - a_{jkl}$ are the values of attribute l for plant life form k in vegetation class i and j

w_{kl} is the weight for attribute l of plant life form k

Changes in threshold values are obtained by computing the Euclidean distance between the original and a modified set of values. This distance measures how far in threshold space the perturbed set of threshold values is from the original one. The distance is calculated by computing the square root of the sum of the differences between original and altered values standardized by the original values (necessary only when the threshold values have different units). The equation to compute the sensitivity values is:

$$\text{Sensitivity} = \frac{1 - \Sigma_k \{\min (V_{ik}, V_{jk}) * [1 - \Sigma_l w_{kl} |a_{ikl} - a_{jkl}|]\}}{\sqrt{\Sigma_n ((\mathbf{p}_{nominal\,n} - \boldsymbol{p}_{varied\,n}) / \boldsymbol{p}_{nominal\,n})^2}}$$

where,

$\mathbf{p}_{nominal}$ is the original threshold value

$\boldsymbol{p}_{varied}$ is the altered threshold value

n is the number of threshold values used

Most likely alternative classification

The goal here is to identify which is the *most likely alternative* classification that may occur when uncertainty in threshold values is considered. The most likely alternative classification is defined as the classification, that being different from the original, results from using the perturbed set of threshold values that is closest to the original set of threshold values. This is accomplished by searching for those threshold values that can change the resulting classification by deviating the least from the best estimates set of values. The distance between the perturbed set of threshold values and the original set is calculated with the Euclidean distance, as indicated above.

While sensitivity measures indicate the rate of change in the output relative to changing inputs, the most sensitive input-output pair is not necessarily the most likely. This is because sensitivity quantifies the difference in vegetation classes (original class versus the one resulting from modifying thresholds) relative to the difference in threshold values while likelihood quantifies a difference in threshold

values regardless of how different the two vegetation classes are. So the most sensitive outcome is the one that results from a small deviation of the original set of threshold values and a large difference in vegetation classes. On the other hand, the most likely outcome is the one that results from a small deviation of the original set of threshold values no matter how different the vegetation classes.

Implementation

A genetic algorithm is used to find the threshold parameter for which maximum sensitivity occurs, and the most likely alternative class in each spatial location. A genetic algorithm is a global combinatorial search methodology based on principles of natural selection (Holland, 1992; Forrest, 1993; Cartwright, 1995; Mitchel, 1996). This technique simulates "evolution" of a solution set, and provides an efficient search of threshold spaces for those combinations of threshold values that result in a maximal sensitivity value or that result in the most likely alternative classification. The search space is defined according to the range of variability of each threshold value. Genetic algorithms simulate a population of individuals. Each individual is composed of a group of genes representing a particular solution. Each gene represents a threshold value and each individual in the population is a particular realization of a set of threshold values. Each individual has a fitness, defined by a computable function described below, which is used to probabilistically select the most fit individuals. Mutation and crossover of genes is used to generate new individuals, representing new, generally improved, combinations of threshold values.

To evaluate the fitness of an individual, a fitness function (also called "objective function") is defined. The fitness function used to find the sensitivity value determines the sensitivity value of the individual's threshold combinations at a particular spatial location by perturbing thresholds at the specified location and computing the sensitivity statistic described above. Using the computed maximum sensitivities at each location, a global maximum sensitivity is determined. The fitness function used to find the most likely alternative classification determines the distance of the individual's threshold combinations that are able to change the output classification at a particular spatial location. This is carried by perturbing thresholds according to their range of variability and, when the output classification changes, computing the distance as described above.

Step 2: Interpretation of Model Results

A measure of confidence in model results was developed to help the modeler judge how model predictions are affected by uncertainty in threshold values. This measure of confidence integrates information about attributes that characterize the output classes with information about their spatial configuration. The specific class attributes to be considered are domain dependent. For applications to MAPSS, we used ecological dissimilarity and the spatial configuration of the vegetation classes. First we determine how "ecologically dissimilar" the output classes are with respect to the most likely alternative output classes. The dissimilarity is measured with the index developed by Sykes et al. (1999) described above. The ecological dissimilarity considers the floristic and structural differences of the vegetation based on attributes of the vegetation classes. Dissimilarity values range from 0 to 100. Similar vegetation

classes are 0 and very different classes are close to 100. If the application does not involve vegetation classes, the index must be adapted to estimate dissimilarity within the context of the specific domain.

Once the dissimilarity measure is established in each spatial location, "core" and "boundary" areas of each class are defined. A core area corresponds to a group of spatial locations of one class that is surrounded by the same class. They correspond to the center or nucleus of a class. A boundary area is considered an intermediate habitat (e.g. an ecotone, Barbour et al., 1987, p 40) indicating a change from one class to another. It is composed of spatial locations surrounding a core area of the same class and limiting with spatial locations of another classes.

Finally, the dissimilarity information is integrated with the spatial information (core vs. boundary area) to create a measure of confidence for the model results (Figure 1). Confidence in the model is very low if the class output from the model in a spatial location is ecologically different from the alternative class and it is located in a core area. Confidence in the model is moderate if the same change in vegetation class occurs in a boundary area. On the other hand, if the output class and the alternative class are very similar and the spatial location is in a core area, then the confidence is classified as moderate. While for the same difference located in a boundary area the confidence is classified as high.

Core areas are expected to be stable and not subject to radical changes when some parameters are varied slightly. Since ecotones are areas of transition, minor differences in vegetation classes are expected. It is important to note that we are

strictly referring to stability in the classification as opposed to biophysical stability (i.e., vegetation composition and structure through space and time), which is out of the scope of this paper.

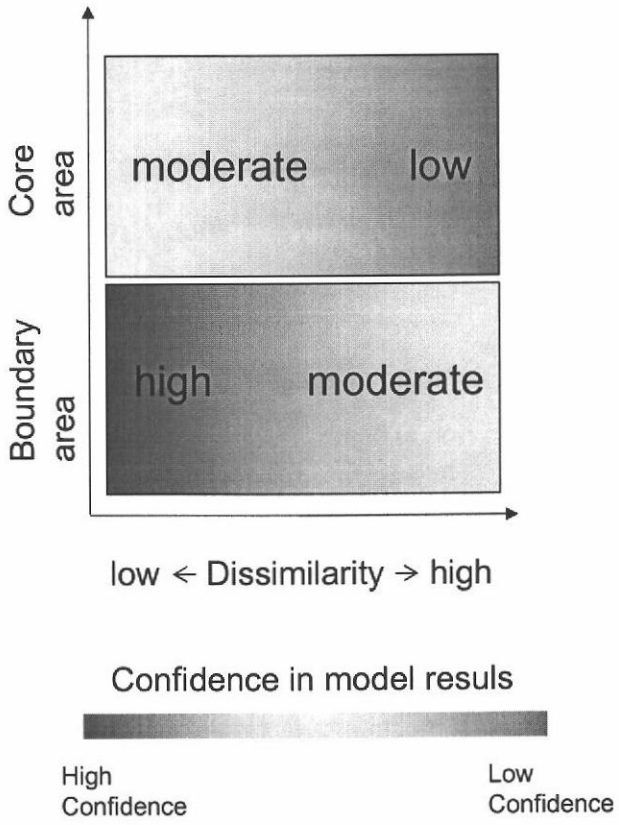


Figure 1. Measure of confidence in model results. The confidence in model results is the lowest when in a core area the dissimilarity is high. On the other hand, the confidence in model results is the highest when in a boundary area the dissimilarity is low. Moderate confidence occurs when the dissimilarity is low in a core area or high in a boundary area.

The measure of confidence in model results does not include information about the maximum sensitivity values, because interpretation of the sensitivity can only be done in a contextual manner. That is, knowledge about model behavior must be integrated with knowledge about the reality being modeled. The measure of sensitivity provides information about model behavior, representing the expected

change in model output when the uncertainty in threshold values is considered. High levels of sensitivity indicate that model predictions are dependent on the threshold values while low levels indicate that changes in threshold do not affect the model results. The measure of sensitivity does not provide any insight into the validity of model behavior. Any level of sensitivity may be desirable if it represents the modeled phenomena. When the correspondence between model sensitivity and reality is not reached, the sensitivity measure suggests a problem in model formulation. It is ultimately the task of the modeler to understand the information provided by this measure and how it represents the real phenomena.

Step 3: Incorporation of Alternative Results into Model Output

The goal of this stage of the analysis is to incorporate the most likely alternative classes into model results. Each spatial location in a landscape can have more than one classification: the original output class from the best estimates set of threshold values and the alternative class, from perturbing the threshold values. The presence of more than one class in a spatial location indicates a model limitation to precisely classify that spatial location. The extent an outcome class can represent the correct classification depends on its level of certainty.

Fuzzy sets (Kruse et al., 1994; Klir and Yuan, 1995; Sangalli, 1998) are used to determine the level of certainty of the predictions. One set is used to represent the certainty of the original output class and another set is used to represent the certainty of the alternative class. Figure 2 illustrates an example of a membership function, considering only one possible alternative class. In this example, the most

extreme scenario would be for an original class to have the same level of certainty (0.5 degree membership) as an alternative class. Information about degree of membership and alternative classes is then graphically incorporated in the results.

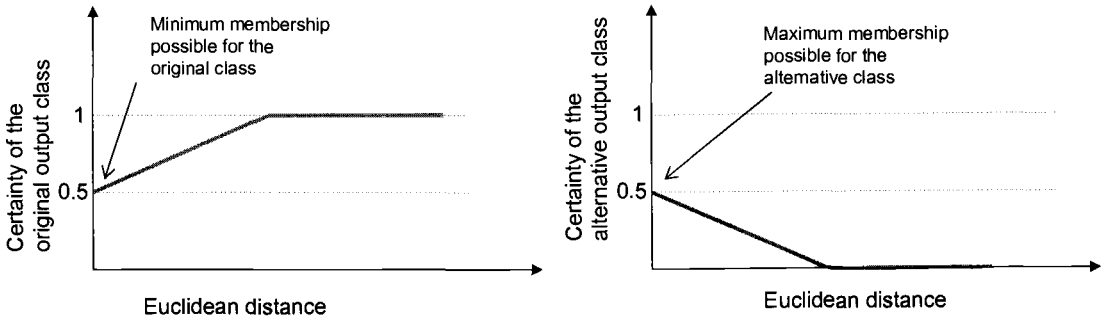


Figure 2. Degree of membership in the original class output from the model and in alternative class. The degree of membership represents the level of certainty of a class. For the original class, the level of certainty is given by the degree of membership in the set that represents the certainty of the original class. For the alternative class, the level of certainty is given by the degree of membership in the set that represents the certainty of the alternative class.

At each spatial location in a landscape, the level of certainty of the original output class is given by the degree of membership in the set that represents the certainty of the original class. If an alternative class exists, its level of certainty is given by the degree of membership in the set that represents the certainty of the alternative class. The degree of membership can take values ranging from 0 to 1. The sum of the degrees of membership through all classes should be equal to 1. The degree of membership is calculated as a function of the Euclidean distance between the original set of threshold values and the set used to determine the most likely alternative class. When the distance is very small, the original output class receives a similar level of certainty as the alternative class. In this case, the model is unable to distinguish between the two classifications (the original model output and the alternative classification). As this distance increases, the level of certainty of the

original output class also increases while the level of certainty of the alternative class decreases. If there is no alternative class then there is complete certainty that the original output class represents the correct classification.

Model Description

Mapped Atmosphere-Plant-Soil System (MAPSS) is a combination of rule-based and process-based biogeographical model that predicts potential climax vegetation at global or regional scales (http://www.fs.fed.us/pnw/corvallis/mdr/mapss). This model is used in the context of global climate change to determine how vegetation distribution varies under different climatic scenarios. The main assumption is that vegetation distribution is constrained by the availability of moisture and energy for growth. MAPSS first calculates the maximum potential leaf area index (LAI) of woody trees or shrubs and grass life forms that can be supported at a site. Then, according to the LAI values classifies the vegetation into 21 possible vegetation classes defined by the Vegetation/Ecosystem Modeling and Analysis Project (VEMAP 1995). The output of the model is discrete and is represented by the spatial distribution of vegetation classes. Although the domain of operation of MAPSS is limited to individual spatial locations with no interactions among them, ecosystem properties (e.g., vegetation distribution) also emerge as large-scale patterns. Model output has a hierarchical structure, where the lower level of the hierarchy contains information that pertains only to individual spatial locations and higher levels contain information that emerges from the grouping of spatial locations (i.e. vegetation patterns).

The vegetation classification is implemented by a set of ecophysiological rules. These rules are based on climatic thresholds, the presence/absence of three lifeforms (trees, shrubs and grasses) and their LAI values. They also take into account different leaf characteristics, thermal affinities and seasonal phenology. Either trees or shrubs are assumed to be dominant in a site, but they are mutually exclusive. Table 1 indicates the threshold values used in MAPSS as well as their function in the model.

Threshold	Value	Function
GDD (°C)	0.0	Ice boundary
	615.0	Tundra boundary
	1165.0	Taiga-tundra boundary
LAI (m^2/m^2)	9.0	Limit the temperate deciduous from the northern hardwood
	3.75	Set the lower limits of the forest
	2.1	LAI above which shrub becomes chaparral
	2.0	Set the limits between mid and tall grass
	1.15	Set the limits between short and mid grass
	0.45	Differentiate semi-desert from short grass
	0.1	Differentiate desert from shrub savanna
	0.1	Differentiate desert from grass

Table 1. This table describes the thresholds used in MAPSS and their function in the model.

For example, a given spatial location is classified as forest if the tree LAI resulting from the competition of water and light is higher than 3.75 (LAI-forest-threshold). According to the phenological and leaf form status and thermal zone, the spatial

location is further classified into one of the nine possible forest classes: *evergreen needle (taiga); evergreen needle (temperate), mixed cool, cool broadleaf, deciduous broadleaf, mixed warm, tropical evergreen broadleaf* and *tropical seasonal forest.*

Case Study

This case study applies the uncertainty analysis methodology to MAPSS, run for a landscape represented by a set of gridded cells each cell having geographical (latitude-longitude) dimensions at a ½ degree resolution. The uncertainty in the threshold value is represented by a uniform distribution that ranges from 20% plus and minus from the best estimates threshold value. The selection of this probability distribution represented the best knowledge available about the threshold values. The area selected for the study is the mountainous region of North East Oregon and neighboring Washington. This is a climatically and topographically heterogeneous area characterized by three main vegetation regions: (1) forest, (2) steppe, and (3) shrub steppe. MAPSS classifies the vegetation of the steppe and shrub regions correctly. However, the forest region is not well captured and it is classified as savanna. Twelve threshold values were used in this analysis; two of them represent GDD values and ten represent LAI values.

Results

The results from the first step of the methodology are depicted in Figure 3a which shows the sensitivity surface for the selected area. This surface is created using the maximum point of sensitivity in each spatial location. Figure 3b shows the alternative

classes. In the sensitivity surface, it is possible to observe that the shrub steppe region of Washington is not affected by changes in threshold values. Its sensitivity is zero and has no possible alternative class. There are areas where the sensitivity is high and for which there are alternative vegetation classes. On the other hand, there are areas where the sensitivity is very low and for which there also are alternative vegetation classes.

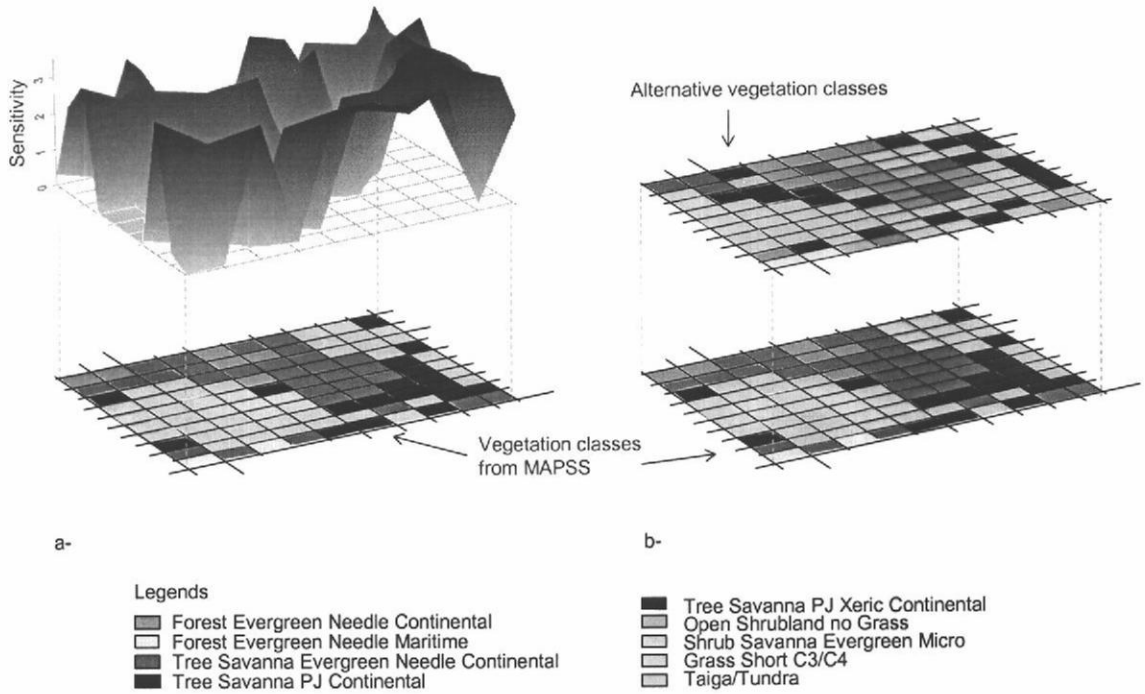
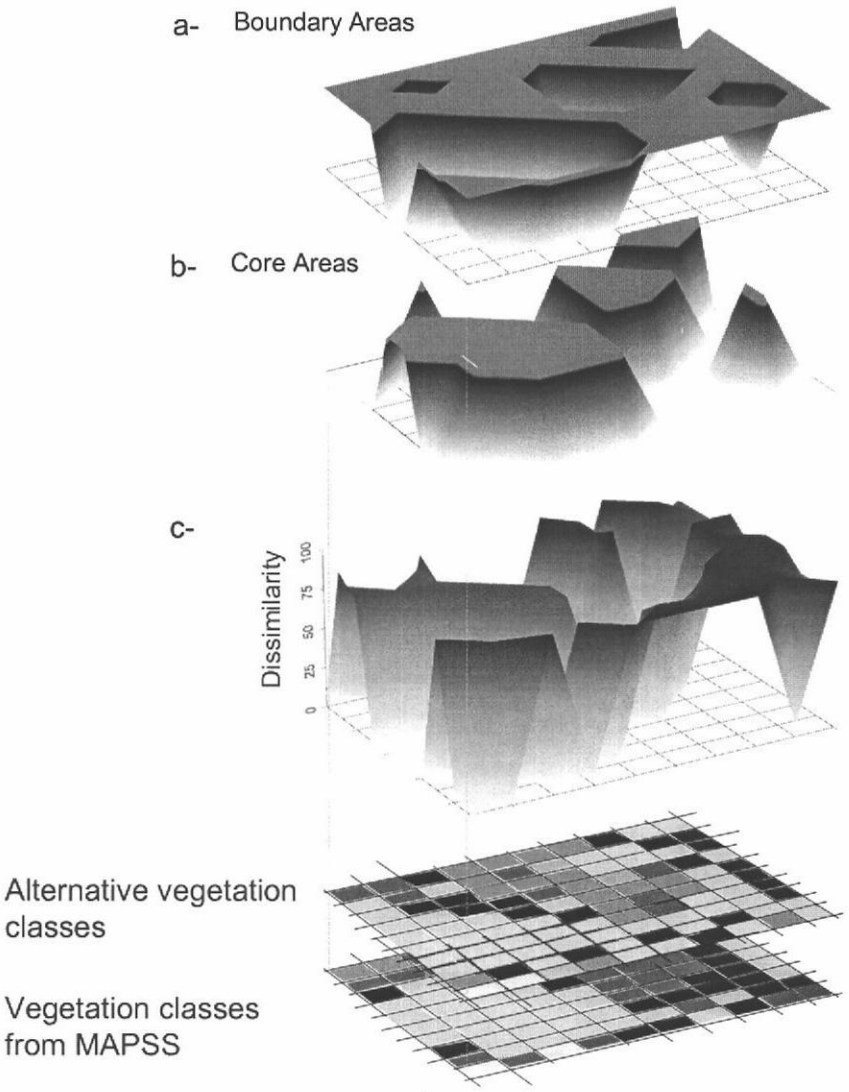


Figure 3 a-b. (a) Maximum sensitivity values. (b) Most possible alternative vegetation classes.

To correctly interpret the effects of the uncertainty, the measure of confidence in model results is computed as suggested in step 2 of the methodology. The results of the dissimilarity index are shown in Figure 4a and the core and boundary areas are depicted in Figure 4b and c. Figure 5 shows the final measure of confidence for the area selected. The shrub steppe area in Washington is classified with "high

confidence". As was previously shown, these spatial locations have very low sensitivity and correspond to a core area with no possible alternative vegetation classes. On the other hand, some of the spatial locations with high sensitivity have a high ecological dissimilarity, but are in boundary areas and so are classified with "moderate confidence". Some spatial locations in core areas have a large ecological distance between the output class and its alternative. These spatial locations are classified with "low confidence" and correspond to locations where the model results must not be trusted.



Figures 4 a-c. (a) Boundary area. (b) Core area. (c) Ecological dissimilarity between the vegetation classes produced by MAPSS and the most possible alternative classes.

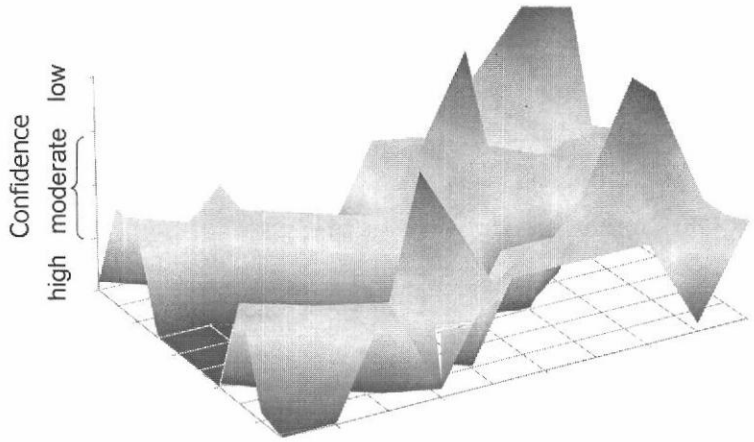


Figure 5. Measure of confidence in model results

Once the levels of sensitivity are computed and alternative vegetation classes found, the information is incorporated into a final analysis. The degrees of membership into the vegetation classes were calculated using the function as described in step 3 and depicted in Figure 1. Figure 6 shows the output classes from MAPSS once the alternative vegetation classes are included. Figure 7 depicts the final results of the methodology, including alternative classes, sensitivity surface, and measure of confidence in model results.

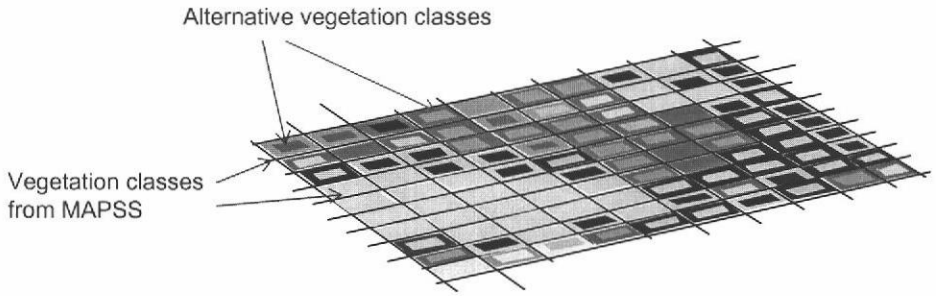


Figure 6. Alternative vegetation classes included as part as model result. When there is no alternative class, a spatial location displays a single color, which represents the vegetation class output from MAPSS. When there is an alternative class, a spatial location displays two colors: the outer color represents the vegetation class output from MAPSS and the inner color represents the alternative class. The size of the alternative class represents the degree of certainty about the classification.

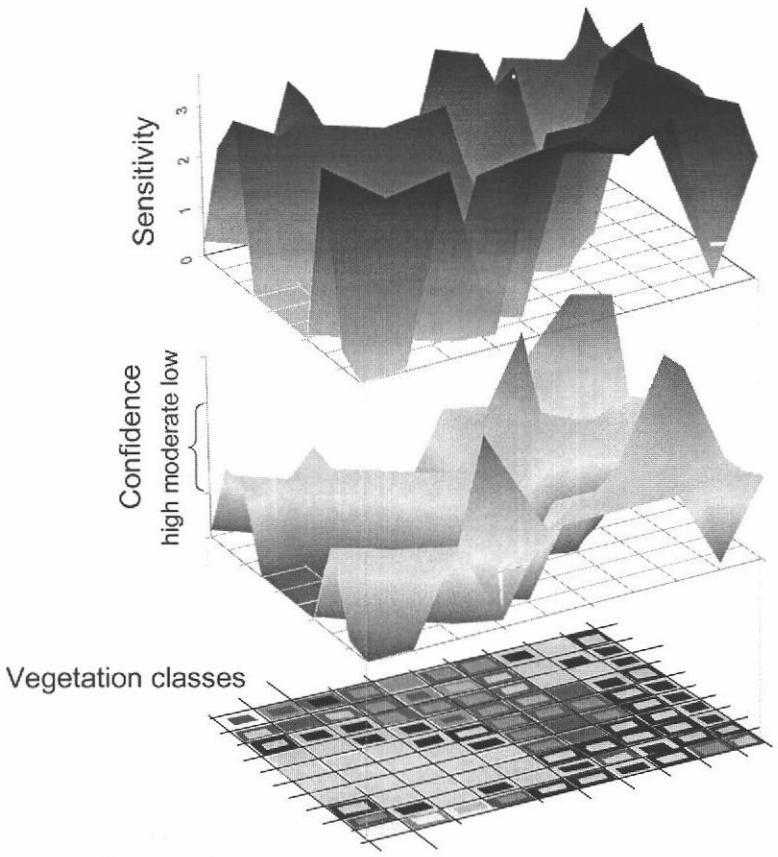


Figure7: Final output. From top to bottom: sensitivity surface, measure of confidence in model results and original and alternative vegetation classes.

Discussion

The methodology presented here has been designed to improve model output evaluation and interpretation by investigating model behavior in threshold regions of rule-based classification models. While this approach is at first glance similar to a Monte Carlo analysis in its goal to determine how the uncertainty propagates through the model and affects its results, the present methodology differs in the way in which it reaches this goal and in the type of information it provides. In a Monte Carlo analysis the focus is to determine the probability distribution of the prediction results. In the methodology presented here the focus is to find those instances in which the output is not affected by threshold values and those in which it is dependent on them (i.e., classification is uncertain). It determines which are the most likely alternative scenarios and also provides the tools to interpret the results of the analysis in an ecologically meaningful way. The goal is to be able to infer ecological significance commensurate with assessed model uncertainty. In a standard Monte Carlo analysis the probability distributions of predicted outcomes is estimated by running the model under different model scenarios. The parameters are considered random variables represented by probability distribution functions. Each scenario is composed of a different set of parameter values generated by random selection from their probability distributions, without discriminating for the effect on the output. In the present methodology, the original set of parameters represents the best estimates and investigates how deviations from these best estimates (i.e., uncertainty) can affect the outcome classification. This approach restricts its focus to find only the parameter values that can generate a highly sensitive state or that can change the resulting classification. Further, the outcome variability in a standard Monte Carlo

analysis is represented by a probability distribution. In contrast, the approach taken employs fuzzy sets to represent outcome uncertainty.

Genetic algorithms, although computationally intensive, proved to be an effective technique to search parameter space and find the most sensitive and most likely outcomes. One advantage of this technique is that the objective function can be easily modified to allow testing different aspects of the model. By using fuzzy sets it was possible to determine the level of certainty associated with the final classification thereby providing the necessary information to graphically represent and incorporate uncertainty in the results. A drawback with this representation is that different certainty functions can lead to different results. For this reason, it may be necessary to investigate which are the most appropriate functions for the problem. For example, in the case study the function that represents the certainty of the original class (Figure 2) has a minimum value of 0.5 that increases linearly until it reaches a value of 1. This function corresponded to the simplest representation of classification certainty. However, another function, with different minimum value or shape (e.g. exponential) could have been selected if more information on the behavior of vegetation classes was on hand.

The results derived from the methodology used here indicate that the combined use of a genetic algorithm and fuzzy sets to assess uncertainty is very useful. By capturing the propagation of threshold value uncertainty, these methods allow estimation of the uncertainty associated with model results and their incorporation as an explicit part of model output. For models like MAPSS, this approach addresses the central issue of interpreting ambiguous model results when the calculated tree,

shrub, or grass LAI values or GDD are in the vicinity of the threshold values. In such cases, even though the classification output favors a particular vegetation class, the output can easily "switch" to another vegetation class. In this scenario, it is not possible to discern which is the most *ecologically meaningful* result because the vegetation patterns generated by the model are highly dependent on the threshold values used, and hence, uncertain and difficult to interpret. This methodology allows incorporation of the most likely classifications as an integral part of the model output and facilitates the evaluation and interpretation of the results through a set of ecologically based metrics. In contrast to most other methods the present approach is designed to identify ambiguous results and provides a quantitative method to asses their significance.

Conclusions

This methodology provides a conceptual framework for evaluating and interpreting modeling results comprehensively where model capabilities and model limitations are both acknowledged. It uses uncertainties as a source of information to determine the scope of model inference, identifying those instances in which the predictions are reliable and those in which they are not. This methodology analyzes model behavior by studying how the model reacts to perturbations in threshold values. It also determines the most likely alternative output classes that can be generated by the model when uncertainty in the threshold values is considered. It further provides the analyst with tools that guide the interpretation of uncertainty effects from an ecological perspective.

In the case study presented here, the ecological interpretation of model results considers the structural and floristic differences between the output vegetation classes and their alternative scenarios as well as the spatial configuration of these classes. In the last step, it incorporates alternative class information into the model results to make the uncertainty in model predictions explicit. The results clearly indicated the areas in the landscape affected by constructs of the rule-based representation. In these areas, the classification was ambiguous and confidence in model results was low (e.g., the core area of the Tree Savanna PJ Xeric Continental with Shrub Savanna Evergreen Micro as the alternative vegetation class). On the other hand, the results also indicate the areas where the confidence in the classification was moderate and the areas where the confidence was high (e.g., the core area of the Shrub Savanna Evergreen Micro is not affected by changes in threshold and do not have alternative class). Overall these results inform about the deficiencies in model predictions (e.g., vegetation classes miss-represented) and how significant these deficiencies are. This information provides the modeler with an element of judgment for the corroboration of the scientific hypothesis embedded in the model as well as for the establishment of the degree of confidence in the output vegetation classification.

From a model evaluation perspective, the information provided by this methodology can be used to identify flaws in model logic and to determine what type of improvements may be needed in a model. From a model interpretation perspective, this methodology provides the scientist and decision maker with valuable information about the quality and limits of model predictions. Overall, this methodology focuses on the significance of uncertainty in ecological models. While in this methodology

only considers uncertainty in threshold values, future research should be directed to investigate other sources of uncertainty that may also affect model output. Finally, this study brings a concrete approach to representing both what the modeler knows and what it is unknown or uncertain.

References

Barbour, M.,G. L.H. Burk and W.D. Pitts. 1987. Terrestrial plant ecology. Second edition. The Benjamin/Cummings Publishing Company Inc., 634 pp.

Bradshaw, G.A. and J. G. Borchers. 2000. Uncertainty as information: narrowing the science-policy gap. Conservation Ecology, 4(1):7. [online] URL://www.consecol.org/vol4/iss1/art7

Cartwright, H.M. 1995. The genetic algorithm in Science. Pestic. Sci., 45:171-178.

Campolongo, F., A. Saltelli, T. Sorensen and S. Tarantola. 2000. Hitchhiker's Guide to sensitivity analysis. In Sensitivity Analysis, Edited by A. Saltelli, K. Chan and E.M.Scott. Willey Series in probability and statistics, pp 15-45.

Forrest, S. 1993. Genetic algorithms. Science, 261:872-878.

Haefner, J. W. 1996. Modeling Biological Systems, principles and applications. Chapman and Hall, 473 pp.

Holland, J.H. 1992. Genetic algorithms. Scientific American, 267 (1): 66-72.

Kann, A. and J.P. Weyant. 2000. Approaches for performing uncertainty analysis in large-scale energy/economic policy models. Environmental Modeling and Assessment, 5:29-46.

Katz, R. W.1999. Techniques for estimating uncertainty in Climate Change scenarios and Impact studies. In: Representing Uncertainty in Climate Change Scenarios and Impact Studies. Proceeding of the EU Concerted Action Initiative ECLAT-2

Workshop 1, Helsinki, Finland, April 16th - 18th, 1999, Climatic Research Unit, Norwich, UK., pp 38-53.

Klir, J.G. and Bo Yuan. 1995. Fuzzy sets and fuzzy logic. Theory and applications. Prentice Hall, 574 pp.

Klir, J.K. and M.J.Wierman. 1999. Uncertainty-Based information. Elements of a generalized information theory. Physica-Verlag, 168 pp.

Kruse, R.,J. Gebhardt and F. Klawonn. 1994. Foundations of fuzzy systems. John Wiley & sons, 265 pp.

Mitchel, M. 1996. An introduction to genetic algorithms. Cambridge, Mass. MIT Press, 205 pp.

Morgan, M. G. and Henrion, M. 1990. Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis, Cambridge University Press, Cambridge, UK, 344 pp.

Neilson, R.P. 1995. A model for predicting continental scale vegetation distribution and water balance. Ecological Applications, 5:362-386.

Omlin, M. and P. Reichert. 1999. A comparison of techniques for the estimation of model uncertainty. Ecological Modelling, 115:45-59.

Sangalli, A. 1998. The importance of being fuzzy and other insights from the border between math and computers. Princeton University Press, 173 pp.

Saltelli, A. 2000. What is Sensitivity Analysis?. In Sensitivity Analysis, Edited by A. Saltelli, K. Chan and E.M.Scott. Willey Series in probability and statistics, pp 3-13.

Sykes, M. T, C. Prentice and F. Laarif. 1999. Quantifying the impact of global climate change on potential natural vegetation. Climate Change, 41: 37-52.

VEMAP Members. 1995. Vegetation/ecosystem modeling and analysis project: comparing biogeography and biogeochemistry models in a continental scale study of terrestrial ecosystem responses to climate change and CO2 doubling. Global Biogeochemical Cycles. 9(4):407-437.

Chapter 3

Process level sensitivity analysis for complex ecological models

Ecological Modelling
P.O. Box 181
1000 AD Amsterdam
The Netherlands

Introduction

The use of process-based models as tools for scientific inquiry is becoming increasingly relevant in ecosystem studies. Process-based models are artificial constructs that simulate an ecosystem by mimicking the functions of its component processes. A process-based approach differs with other modeling approaches in several ways. Generally, process-based models (e.g., 3PG, Landsberg and Waring 1997; BIOMASS, McMurtrie 1985) are contrasted with other representations which are formulated using empirically derived relationships and rules (e.g., ORGANON, Hann et al. 1997). Perhaps most importantly, process-based modeling provides a framework with the necessary flexibility to simulate the range of behaviors which may vary across scale and biophysical conditions that other modeling approaches lack.

Structurally, process-based models are characterized in terms of their component processes and their inter-relationships which represent biophysical phenomena. Much of the dynamics of the model which emerge are not explicitly encoded but rather evolve over time with iterative interactions. Because the modeler is responsible for creating the model as well establishing that the behavior generated is appropriate and semantically equivalent to the phenomena being modeled, s/he must also understand the nature and source of this emergent behavior. Assuring a model meets its specifications requires taking into account the many sources of uncertainty in the modeled processes and relationships. Uncertainty refers to what it is unknown about the ecosystem modeled and it can impose severe limits on the ability to make inference and the extent of this inference. In other words, good modeling practice depends not only on understanding how well the model mimics

system behavior but how well the modeler understands the model's intrinsic uncertainty. Building a model without equal rigor of uncertainty assessment is incomplete. Uncertainty is integral to modeling.

Uncertainty arises from a deficiency in information (Klir and Wierman 1999) and it typically is reflected in a lack of knowledge about the exact parameter values, the data fed into a model and the process specification. There are several extant techniques available for use to characterize and understand the effects of uncertainty (Gardner et al. 1990, Oderwald and Hans 1990, 1993, Reynolds and Ford 1999, Katz 1999, Omlin and Reichert 1999, Campolongo et al. 2000, Green 2000, Kann and Weyant 2000, Parysow et al. 2000). The most popular methods are classified as either sensitivity or uncertainty analyses. Sensitivity analysis is a general approach to study model behavior (Haefner 1996, Elston 1992) where the central aim is to determine the rate of change in model output as a function of variability in input factors (R.W.Katz, 1999). This variability typically represents the uncertainty associated with the parameters or the data fed to a model. The different sensitivity analysis techniques have also been expanded to include uncertainty in model structure and assumptions. The measure of model's sensitivity indicates how much an output can vary relatively to the variations in these input factors. High sensitivity indicates that a small change in the input data or parameters values can generate a relatively large change in model results. Conversely, low sensitivity indicates that even large changes in input data or parameter values do not significantly affect model outcome. As defined by Saltelli et al. (2000), sensitivity analysis studies the relationship between information flowing in and out of the model. In contrast, uncertainty analysis measures the uncertainty of model's results. This analysis is

concerned with estimating the overall uncertainty of model output given the uncertainty associated with parameters or input data (Saltelli et al. 2000). Despite the availability of techniques they present some difficulties when used to analyze complex ecological models. From an operational point of view, the application of these techniques may constitute a challenging task due to the large parameter space and the high degree on non-linearity these models present. From a conceptual point of view, difficulties may arise when the analysis does not parallel the way in which a model is thought; that is in terms of processes that interact among each other to generate a realistic representation of how the ecosystem works. This is important because understanding how the output is generated constitutes an essential part of the corroboration of a model and the assurance of its quality (Beck 2002). Also, ignoring a model is conceptualized and organized in terms of processes hinders the investigation of the effect process uncertainty may have in model behavior. Generally, process-level uncertainty is "black-boxed" and evaluation of model uncertainty is restricted to investigating explaining model behavior by analyzing the relationships between input-output pair patterns. Thus, uncertainty intrinsic to ecosystem process conceptualization is frequently avoided. Any variation derived from the process level is subsumed and confounded in the input-output analysis.

Process uncertainty is a multidimensional concept that has many forms, sources, and attributes (Zimmermann, H.-J. 2000, Gardner et al. 1990). Process-level uncertainty is contrasted with uncertainty derived from variation in input factors, such as parameters or input data. Process uncertainty represents the modeler's knowledge, or lack of knowledge, about the modeled system. This uncertainty can

derive from a mis-conceptualization, or an insufficient understanding, of the system behavior. For example, one source of process uncertainty is the lack of theoretical knowledge, and lack of consensus about how a process operates (e.g. how plants allocate carbon within an ecosystem). Another source is the lack of knowledge about how processes interact with each other (e.g., how individual cycles relate to each other and how they interchange information across scales and conditions). Yet another source is an insufficient understanding about the behavior and magnitude of parameter values which govern one or more sets of biophysical processes.

Here, we propose a methodology of sensitivity analysis that is designed to investigate the effects of process-level uncertainty on model dynamics. In this approach, we extend the meaning of sensitivity analysis to the process level. By extending modeling investigation to a higher-order level of model behavior and construction; this higher-order analysis is therefore able to capture the uncertainty which is introduced and propagates to the network of model processes. Process-level sensitivity analyses are designed to provide a more comprehensive assessment for assuring that the behavior generated by the model is appropriate and semantically equivalent to the phenomena being modeled. The methodology, therefore, is not only intended to provide a description of model sensitivity but also to help the modeler probe her/his knowledge of the theory that was used to design and is embedded in the model. Such a shift to a higher-order also has the technical advantage of helping reduce the necessary search space dimensionality which can be computational intensive and limiting in many models. Further, this type of approach seeks to bring more rigor to the inference-making capacity of the modeler by closing the uncertainty gaps between various overlapping (e.g., distributed

parameters across levels and sub-models) or hierarchical scales (e.g., input to process-level) and domains (e.g., research, management and policy; Bradshaw and Borchers, 1999; Funtowicz and Ravetz, 1996). These are tools for evaluating integrated and complex systems.

Methods

We begin by describing a general approach for building a process-level sensitivity analysis. More specifically, we investigate the impact that varying one or more processes may have on overall model behavior. The sensitivity analysis effects a perturbation of the process or processes themselves and study how this perturbation affects overall model behavior and its output. The focus of this approach is in analyzing model behavior in terms of its processes and interactions, considering that a good modeling practice requires understanding model dynamics when uncertainties in processes are considered. This is different than a traditional sensitivity analysis that focuses in understanding how the model reacts to perturbations in parameter or input data values. Two aspects of model behavior are analyzed: (1) the alteration in process to process interactions which occurs when the model is perturbed, and (2) the variation in model output generated in response to these perturbations. The sensitivity analysis can be described as a succession of three distinct steps (Figure 1).

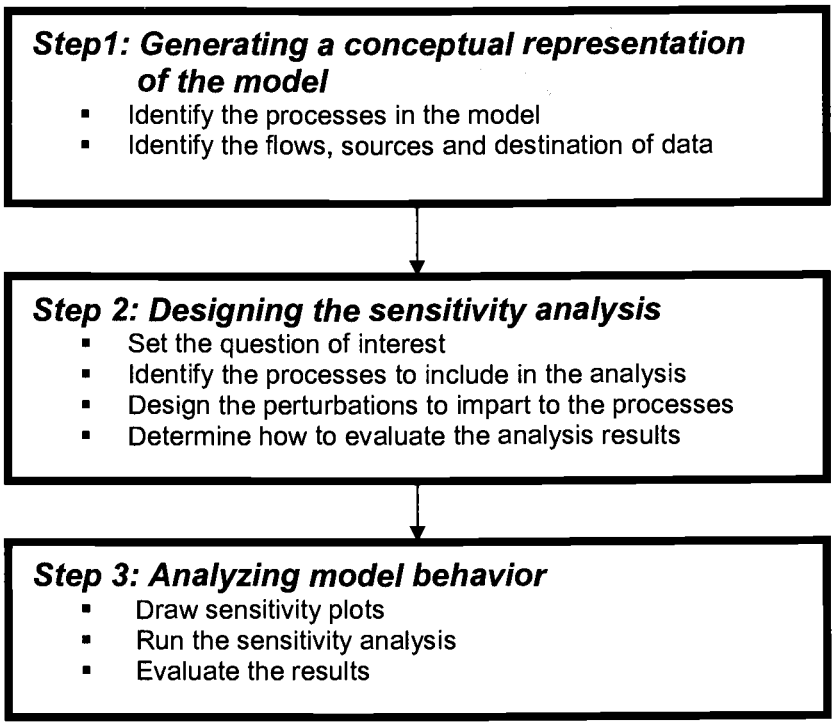


Figure 1. Steps suggested for the application of the methodology.

(1) *Step 1: Generating a conceptual representation of the model*

Generating a conceptual representation of the model consists of representing *what* the model does at the level of a given process, without consideration of its actual implementation (i.e. *how* the model does it). We accomplish this step by conceptualizing the model as a network of processes connected by flows of data. In this conceptualization the individual processes represent what the model components do and the data flows represent the information that is exchanged between processes. The conceptualization can be rendered at multiple levels depending on the modeler's objectives. Thus, the modeler must first decide which set of processes will be targeted in the sensitivity analysis.

1.1. Identify Model Processes. We define process as a functional unit of the model that performs a specific action (e.g., in a vegetation model processes can be plant

functions such as photosynthesis, carbon allocation, maintenance respiration, etc.). Processes represent the transformations performed on the data; that is, a process operates by receiving an input flow of data, changing it and producing an output flow of data, which then becomes input data for another process. For example, *photosynthesis* is a process that uses information about leaf nitrogen content, stomata conductance, photosynthetically active radiation (PAR) and carbon dioxide (CO2) concentration (i.e., input flows of data) to compute the amount of carbon assimilated by plants (i.e., output flow of data). The input data transformed by the process becomes the output data which in turn is utilized by another process as input. Using the previous example, assimilated carbon which forms the output of the photosynthetic process becomes the input in the allocation process which distributes the carbon into different plant compartments.

Processes can be conceptualized and identified at different levels of detail. A process at one level can be decomposed into several sub-processes which represent the process in a higher level of detail. Each level shows a more complete description of what the model does (Fishwick, 1988). For example, carbon cycling through a forest can be considered both as a single process, but also at a more detailed level of description where it is decomposed into a set of sub-processes such as *photosynthesis*, *maintenance respiration*, *growth respiration*, *radiation uptake*, etc. The necessary and desired level of detail and process aggregation are subjective decisions that the modeler must make (see 2.2 below).

1.2. Identify data flows, sources and destinations. Data flow represents movement of data through the model. Data can derive internally by moving from one process to

the other (e.g., carbon assimilated that flows between photosynthesis –source- and allocation –destination-process) or it can derive externally as an input from outside the model (e.g., input radiation). In all cases, however, the task of generating a conceptual representation requires identifying processes, data flows and sources, and destinations of the data within each level of description.

(2) Step 2: Designing the sensitivity analysis

Sensitivity analysis design entails determining how the analysis is to be carried out in accordance with the study objectives. First the modeler must establish the questions the analysis is meant to answer. Then the modeler needs to determine how to go about getting these answers (i.e., design the sensitivity analysis). The design phase involves the following activities: (1) setting the questions of the analysis, (2) identifying process level of detail, (3) designing the types of perturbation that will be exerted on the processes, and (4) determining how to evaluate analysis results.

2.1. Setting the question of interest. During this step it is necessary to establish the questions and scope that the analysis seeks to address. It is based on this question that the investigator designs the analysis, determining how the methodology will be executed and which type of information will be obtained. For example, the investigator may want to focus on the individual impact of a process on some aspects of the model (e.g., which is the impact that changing the allocation process has on the carbon cycle processes and their outcome), or may want to evaluate the simultaneous impact of a group of processes (e.g., which is the impact that changing

the allocation, photosynthesis and respiration processes has on the carbon dynamics).

2.2. Identifying Process Level of Detail. As mentioned in section 1.1 above, the level of detail selected for analysis depends on a number of factors including the various study objectives, confidence of the modeler with regard to the validity of model components and logistics which can constrain data collection and computation capacity. Selecting the appropriate level of analysis directly depends on the perceived need for uncertainty assessment. Process descriptions at higher levels of detail can facilitate the understanding of those aspects of the model that are relevant to the objective of the study. Descriptions at a lower level of detail can be used to encapsulate information of those parts of the model that are not the main focus of the analysis. The final product of this step is the development of a flow diagram that shows the conceptual representation of the model. For example, an analysis that aims to understand the carbon dynamics would require a high level description of the carbon cycle processes because these processes constitute what the modeler wants to study. Conversely, other processes, such as those that pertain to the water cycle, can be described at a low level of detail since they are not the main focus of analysis.

2.3. Designing the perturbation. A perturbation is defined here as a change imparted to a process. A process is perturbed to evaluate the impact that alterations in it has on model behavior. Like the choice of process level, the selection of the magnitude and nature of the perturbations depend on the goals of the analysis and the availability of information. Different types of perturbations can be generated, each

designed to achieve a specific purpose and each appropriate to the given process. Processes can be perturbed one at a time while all the other processes are kept fixed or can be varied simultaneously. The first case studies the individual impact of one process on a particular output variable, while the second case studies the impact of individual factors while other factors are varied as well.

One type of perturbation consists of varying the parameters associated with each process. Even though this type of perturbation manipulates parameters, these parameters are only used as a vehicle to exert a change in a process and they are not subject to any analysis. Here, the main goal is to evaluate how changing a process or set of processes affect model behavior and it is not to measure the contribution of these parameters to the changes. For example, quantum efficiency could be a parameter to be modified in a process that calculates gross primary production (GPP) as a function of incident photosynthetically active radiation; perturbing this parameter generates an increase or decrease in the resulting GPP and provides information to assess the effects that variations (i.e., increases or decreases) in the GPP calculation process have on model dynamics.

When perturbations are effected by altering the parameters, it is necessary to assert that the parameters modified affect only the process of interest. If they affect more than one process, the modification could produce a compound result. Each parameter value can be varied according to the probability distribution that represents the uncertainty associated with the parameter.

Another type of perturbation entails modifying the process formulation itself. The goal here, in contrast with the example above, is to evaluate how different formulations of the given process affect overall model behavior. For example, the modeler may examine GPP as a function of incident photosynthetically active radiation, and contrast this resultant model behavior with the formulation of GPP as a function of incident photosynthetically active radiation but now including the effects of nutrient and water deficits as constraints.

2.4. *Determining how to evaluate analysis results.* This step consists of designing the measurement of the perturbation impact on model behavior. Campolongo et al. (2000) suggest a series of approaches that facilitate analysis result evaluation. In this paper, we determine the range of model results, that is, the relative change that occurred for each input-output pair in a process and the sensitivity of the output data to variations in a process. While the first measure can be used to evaluate whether or not the range in the output is a sound one, the second measure can be used to determine the influence a process has on model output. The sensitivity measure is a linear estimate of the number of units changed in output data (i.e. the response variable) per unit of change in the results of the perturbed process as measured by the output of the process perturbed. When sensitivity has a value of 1 it means a unit of change in a process imparts a unit of change in the in the response variable analyzed. Values that are higher than one imply the response variable is sensitivity to changes in the process while values that are smaller than one imply low sensitivity.

$$Sensitivity = \frac{[(O_p - O_b)/O_b]}{[(P_p - P_b)/P_b]} \qquad \text{(Equation 1)}$$

where,

P_b baseline output from the process perturbed

P_p output from process after perturbing it

O_b baseline response variable that is subject to analysis

O_p perturbed response variable that is subject to analysis

Step 3: Analyzing model behavior

3.1. Drawing sensitivity plots.

To visualize the effects of perturbations in model behavior we construct sensitivity plots. A two dimensional sensitivity plot is created for each input-output pair in a process. Here output data is associated with the vertical axis and the input data is associated with the horizontal axis. The plots show the relative change occurred for each input-output pair in the perturbation of a given process. The relative change is computed using average values of the data. Each point in the graph corresponds to a model run. The sensitivity plots are superimposed into the flow diagrams previously built, to formulate a complete description of model behavior and facilitate the visualization of the results of the analysis (see Figure 2.)

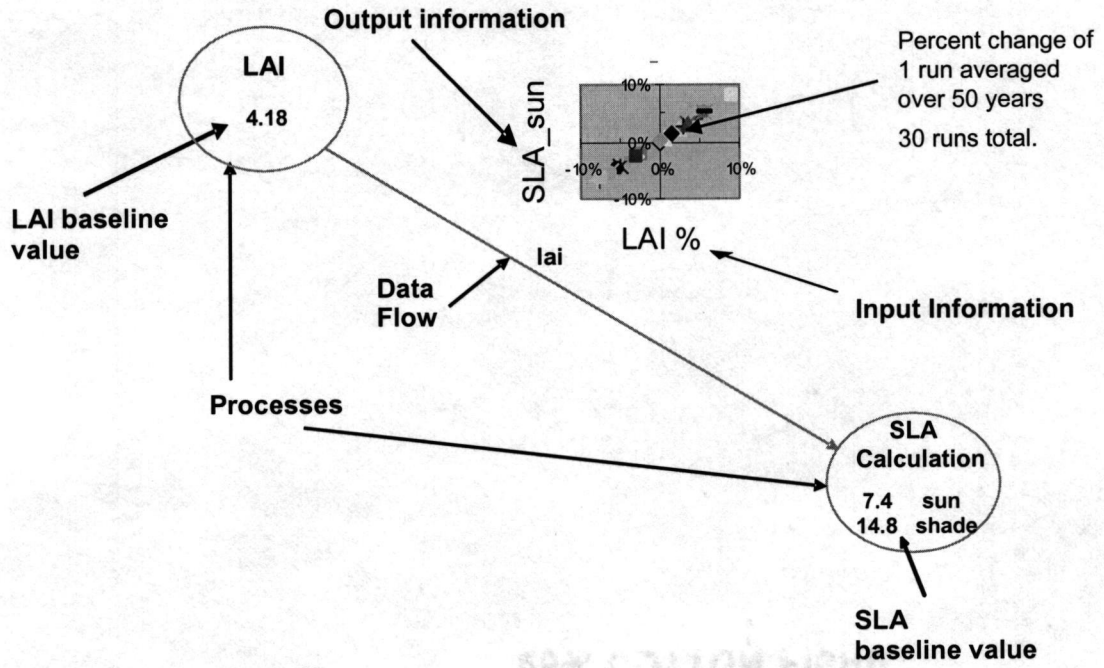


Figure 2. Sensitivity plot.

3.2. Running the sensitivity analysis. In this step, the previously designed perturbations are applied. For each perturbation the model is run to obtain a sample of the model output of interest. The results of the analysis are depicted in the flow diagram with the sensitivity plot superimposed. Figure 2 shows a partial flow diagram with a sensitivity plot. When perturbing a process through its parameters information about the response of the model with respect to changes in parameters can also be gathered.

3.3. *Evaluating the results.* During this step the results of the analysis are examined to assess how the perturbations imparted to a process or group of processes alter the behavior of the model. On one hand, it is necessary to evaluate how the interactions between processes change in response to perturbations and so ensure

that the model resembles the ecosystem being modeled. To that end, it is also necessary to corroborate that the range of model predictions is a sound one.

On the other hand, it is necessary to evaluate the sensitivity of the output to the perturbations in processes. An output that is highly sensitive to a process means that it is strongly dependent on that process, which implies that uncertainties associated with that process may have a large impact on the output. When sensitivity values are constant throughout all the runs that belong to the same perturbation, it means that a change in a particular process always produces a change in the output that is proportional to the former. Sensitivity values that show linear or non-linear trends with respect to a change in a process indicate that variations in the output depend on the magnitude of the change in the process. Sensitivity values equal to zero indicate that changing the process do not produce any effect in the output analyzed.

Data flow diagrams with the sensitivity plots superimposed are used to facilitate this task. When perturbing one process at a time it is convenient to use the perturbed process as a starting point and then, following the direction of the data flows, examine the sensitivity plots found. Analyzing the model in this manner allows an understanding of how the effects of a particular perturbation travel through the model and affect other processes. While in some cases the effect of perturbing a process may be of little influence to other processes, in other cases perturbations may influence the behavior of many processes.

Case study

This case study illustrates the application of the sensitivity analysis methodology to the model Biome-BGC version 4.1.1 (Thornton 1998, www.forestry.umt.edu/ntsg). Biome-BGC is a process-based model that simulates the biogeochemical cycles across terrestrial ecosystems in various geographic regions. The model estimates the daily flow and storage of carbon, nitrogen and water for the vegetation, the litter and the soil components of the ecosystems. The example developed focuses on understanding how variations in the *maintenance respiration* and *leaf area index* (LAI) calculation processes can affect the behavior of the carbon cycle. In particular, how these variations alter the results of the *photosynthesis* process for a site that is located on the coast of the Pacific Northwest of United States. Note, that the purpose of this case study is only to illustrate process-based sensitivity analyses, and by no means intends to be a detailed sensitivity analysis of Biome-BGC. It merely serves as a concrete example of how this methodology might be applied in a real case model and analysis.

This site is located in the Coast Range of Oregon. The Coast Range is a moderately high range with most ridge tops within 450-750 m of altitude, characterized by steep slopes and soils with little development. Our study site, located in the north coast of Oregon has an annual precipitation of 2500 mm, with dry summers and wet winters, and a mean annual temperature of 10 °C. The dominant vegetation is an evergreen needle forest dominated by western red cedar (*Tsuga heterophylla*) and Douglas-fir (*Pseudotsuga menziesii*).

Climatic data developed by the Vegetation/Ecosystem Modeling and Analysis Project Phase 1 (VEMAP1) were used in this example. The data consist of 1 year of daily time series of vapor pressure deficit, maximum and minimum temperature, short wave radiation and precipitation. Using the same climatic year repeatedly, the model was spun up until it reached steady state; this state was used as the initial condition to study model behavior for a period of 50 years (Table 1).

Leaf Area Index (LAI, dimensionless)	*Leaf Carbon (kgC/m2)*	*Leaf Nitrogen (kgN/m2)*	*Available Soil Mineral Nitrogen (kgN/m2)*	*Input of Nitrogen by deposition (KgN/m2/year)*	*Soil Water Potential (MPa)*
4.2	*0.35*	*0.008*	*0.00005*	*0.0004*	*-0.0033*

Table 1. Leaf and soil initial conditions.

Four different boundary conditions of water and nitrogen availability were considered in this example after the model was spun up: (1) water unlimited and nitrogen limited, (2) water unlimited and nitrogen unlimited, (3) water limited and nitrogen limited and (4) water limited and nitrogen unlimited. Water unlimited conditions were reached by increasing the daily amount of precipitation of the site by a factor of 10. Nitrogen unlimited conditions were reached by providing the plants with all the nitrogen they demanded regardless of how much nitrogen was available.

Results

Generating a conceptual representation of the model

Figure 3 shows the conceptual representation of Biome-BGC at different levels of detail. Figure 3*a* depicts the conceptualization at the lowest possible level of detail. The model is represented by only one process, which receives parameters and climatic information as inputs, transforms them and produces outputs such as LAI, net primary productivity (NPP) and GPP. Figure 3*b* presents a more detailed description of the model, where the process shown in figure 3a is decomposed into *carbon*, *nitrogen* and *water cycle* processes and their interactions. The *carbon cycle* process sends LAI information to the *water cycle* process, which, after simulating the water balance, sends back stomata conductance information. The *carbon cycle* also sends information about nitrogen demand to the *nitrogen cycle* process, which, after computing nitrogen balance, returns information about the amount of nitrogen taken by the plants. Table 2 indicates the names and definitions of the inputs and outputs considered in the flow diagrams.

Figure 3c shows a highly detailed description of the model, where each of the three processes in figure 3*b* is further decomposed into several subprocesses. This diagram includes the interactions shown in figure 3*b* with the difference that the increased level of detail in the process descriptions allows a more exact description of the origin and destination of the data flows. For example, figure 3*b* shows that LAI data travels from the *carbon* to the *nitrogen cycle* processes. In figure 3c the flow diagram shows with more detail that it is the *LAI calculation* process, after receiving

leaf carbon as input, the one that computes the LAI that is sent to the *evaporation* and co*nductance* processes in the water cycle.

Variable	**Definition**
allocC	*Carbon allocated (kgC/m2)*
aPAR	*PAR absorbed by canopy (W/m2)*
availableC	*Available carbon (kgC/m2)*
Actual Gross Psn	*Actual gross photosynthesis (KgC/m2/day)*
Assimilation Rate	*Instantaneous assimilation rate (umol/m2/sec)*
C	*Carbon*
DLMR_shade	*Day time maintenance respiration rate* per unit projected leaf area in the shaded portions of the canopy for daytime *(umolC/m2/s).*
DLMR_sun	*Day time maintenance respiration rate* per unit projected leaf area in the sunlit portions of the canopy for daytime *(umolC/m2/s).*
g	*Conductance to CO2 (umol CO2/m2/s/Pa)*
GPP	*Gross primary Productivity(kgC/m2/year)*
LAI	*Leaf area index (dimensionless)*
Leaf C	*Leaf carbon (kgC/m2)*
Leaf N	*Leaf nitrogen (kgN/m2)*
Lnc	*Leaf nitrogen content (kgN/m2)*
m_psi	*Conductance control from soil water potential (dimensionless)*
MR	*Total maintenance respiration (kgC/m2)*
N	*Nitrogen*
NPP	*Net Primary Productivity (kgC/m2/year)*
ppfd	*PAR flux density (umol/m2/sec)*
Potential Gross Psn	*Potential photosynthesis (KgC/m2/day)*
PAR	*Photosynthetic active radiation (W/m2)*
SLA	*Specific leaf area (m2/ kgC)*

Table 2. Name and definition of the input and output data from processes.

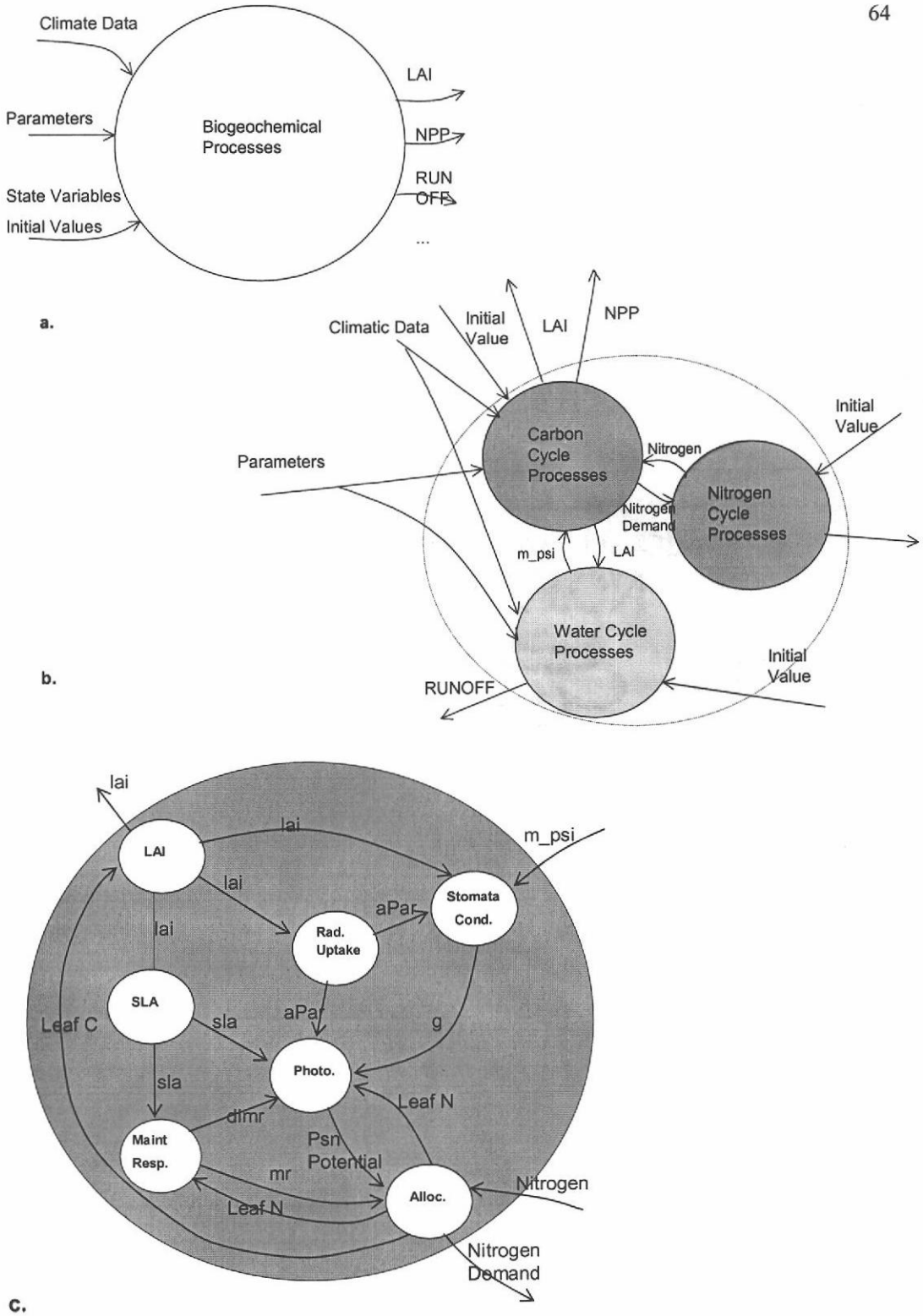


Figure 3. Conceptual representation of BIOME_BGC at different levels of abstraction. (a) conceptualization at the lowest possible level of detail; (b) detailed description of the model, where the process shown in figure 3a is broken up into carbon, nitrogen and water cycle processes; (c) highly detailed description of the model, where each of the three processes in figure 3b is broken up into several processes.

Designing the sensitivity analysis

Setting the question of interest. The goal of the present analysis is to evaluate the effects that variations in *maintenance respiration* and *LAI calculation* processes may have on photosynthesis under different boundary conditions of water and nitrogen availability. Specifically, the aim is to understand how individual variations in these two processes impact the internal working of the model and the following photosynthesis results: instantaneous assimilation rate, potential daily photosynthesis, actual daily photosynthesis and NPP.

Instantaneous assimilation rate is the amount of CO2 assimilated per second by the plant. This rate is used to estimate the potential gross photosynthesis, by multiplying it by the projected leaf area and the day length. The potential gross photosynthesis specifies the amount of carbon the plants can potentially accumulate as a result of photosynthesis on a daily basis. This potential is then adjusted to nitrogen limitations to estimate the daily actual gross photosynthesis. On the other hand NPP measures the final net productivity of the plants on a yearly basis.

Identifying the processes to include in the analysis. When applying the sensitivity analysis methodology to Biome-BGC, seven different processes where identified as relevant to understanding the effects of LAI calculation and maintenance respiration on photosynthesis. The processes included in the analysis were: *LAI calculation*, *specific leaf area (SLA) calculation*, *maintenance respiration*, *stomatal conductance*, *allocation*, *radiation uptake* and *photosynthesis*. The influence of the water and nitrogen cycle were considered through the effects the water cycle processes have

on conductance to CO2 and the nitrogen cycle processes have on the actual gross photosynthesis. Detailed descriptions of these cycles were not included explicitly since they were not the main focus of the analysis. As established in the objectives of the study only *LAI calculation* and *maintenance respiration* process were perturbed.

Designing the perturbation to impart to the processes. The perturbations applied to the processes consist of varying simultaneously the parameters associated with each process. Each parameter value was varied plus and minus ten percent of its value. Uniform random variations of parameters within this range were generated. The perturbations were imparted to one process at a time. While one process was perturbed all the other processes were kept fixed. The processes perturbed were: *LAI calculation* and *maintenance respiration*.

Perturbations to the *LAI calculation* process were generated by randomly changing the parameter that represents the average projected specific leaf area. This parameter is multiplied by the leaf carbon to compute LAI. Changes in this parameter affect the resulting LAI by either increasing or decreasing the area that is projected on the ground per leaf mass of carbon. Perturbations to the *maintenance respiration* process were generated by randomly modifying two parameters: "mpern" that represents the mass of carbon respired per mass of nitrogen in tissue per day at 20 °C and "q10" that is the respiration quotient. Changes in these parameters either increase or decrease the resulting life form's maintenance respiration.

Determining how to evaluate the results of the analysis. To measure the perturbation impacts on the output, three statistics were computed: (1) the range of output variations, (2) the relative change in the output with respect to its baseline value and (3) a sensitivity measure (equation 1). To evaluate how the perturbations affected the interactions among processes in the model, relative changes in each input-output pair process was computed and the results were depicted using sensitivity plots superimposed on flow diagrams.

Analyzing model behavior

Two aspects of model behavior are analyzed: (1) the alteration of process interactions which occurs when the model is perturbed, and (2) the variation of the output data generated by the model in response to these perturbations.

LAI process effects:

LAI is the total projected leaf area per unit area of ground. The LAI process calculates the LAI for the sunlit and shaded canopy fractions. The LAI is used by: (1) the *photosynthesis* process to compute the potential daily canopy photosynthesis, (2) the *SLA calculation* process to compute the SLA for the sunlit and shaded canopy fractions, (3) the *radiation uptake* process to compute how much short wave (SW, W/m2) and photosynthetically active radiation the canopies absorbed (aPAR, W/m2) as well as the PAR flux density (ppfd, umol/m2/s) for sunlit and shaded canopy fractions that indicates how much PAR is absorbed per unit projected leaf area in the respective canopies, and (4) the *stomatal conductance* process to compute the

conductance to evaporated water vapor and CO_2. Even though the perturbations applied to the *LAI* process were the same in all four scenarios, the model responded differently to these changes affecting they way in which processes interacted as well as the output data generated. It was in the scenario where water was limited and nitrogen unlimited where the output from photosynthesis expressed the highest variability in response to changes in LAI.

Alteration in process-process interactions

In all four scenarios variations in LAI produced changes in SLA that were in the range of +/- 9 % of the SLA baseline value. As expected, an increase in LAI produced an increase in SLA in the sunlit and shaded portions of the canopy, with a corresponding decrease in the instantaneous assimilation rate. This is because higher SLA values resulted in smaller concentrations of carbon and nitrogen per unit leaf area and reduced chlorophyll pigments in the Rubisco enzyme for photosynthesis. Increases in SLA were in turn associated with decreases in the rate of maintenance respiration per unit projected leaf area in the sunlit and shaded portions of the canopy for daytime (DLMR_sun and DLMR_shade), which values varied between +/- 10 % of the DLMR_sun and DLMR_shade baseline value.

Variations in LAI also affected the amount of photosynthetically active radiation and short wave absorbed (aPAR and SW) by the canopy. Yet, these changes were very small, being less than 2 % in the scenarios where water was abundant and less than 1 % in the scenarios where water was limited. Increasing LAI increased the amount of aPAR and SW absorbed by the canopy. This increase in aPAR resulted in a

decrease in ppfd, which in turn was associated with a decrease in the instantaneous assimilation rate. Additionally the variations in ppfd affected stomatal conductance to water vapor and CO2, which responded by closing when the levels of ppfd diminished.

Furthermore changes in LAI affected the amount of water evapo-transpired. As expected, increases in LAI were associated with increases in canopy evaporation, changes that in turn affected all other process in the water balance. In all cases increases in LAI were associated with a decrease in soil water potential (psi) values. As water in the soil was depleted, leaf conductance to water vapor and to CO_2 diminished. Final values of conductance varied the most in the scenarios where water was limited, variations that were in the range of +/-9 %. The smaller variations happened in the scenario where water was abundant and nitrogen limited. Here the sunlit conductance remained invariant while the conductance in sunshade part showed a change that ranged +/- 6 %.

Changes in conductance to CO2, ppfd and SLA directly affected the calculation of the instantaneous assimilation rate, which diminished in response to lesser amounts of CO2, ppfd and leaf nitrogen concentration. Even though LAI perturbations produced similar qualitative effects on the instantaneous assimilation rates, they were quantitatively different. It was in the two scenarios, where water was abundant that the highest instantaneous assimilation rates happened; while the smallest values were found where water was a limiting factor and nitrogen unlimited (see the graphs that show the absolute values in Table 3).

The potential gross photosynthesis responded similarly to changes in LAI, where increases in LAI were associated with decreases in the potential gross photosynthesis. However this trend was reversed in the scenario where water was abundant and nitrogen unlimited, since increases in LAI were associated with increases in the potential canopy photosynthesis (see Table 4).

The actual gross and net photosynthesis, was not only influenced by the assimilation rates, but also by the potential daily canopy gross photosynthesis and the nitrogen and water limitations in the system. In the scenarios where the nitrogen was limited, there was not enough nitrogen to fulfill the plant demand (from soil and retranslocation) and in consequence, the amount of nitrogen taken by the plants was much lower than the nitrogen needed. To account for this nitrogen limitation the plants had to adjust the carbon assimilation flux to the availability of nitrogen. Under nitrogen limitation conditions, even when the potential photosynthesis was high, only a limited amount of carbon gained could be allocated to the plants.

From the two scenarios that were nitrogen limited, the one that was also water limited resulted in highest values of actual gross photosynthesis, regardless of having the lowest instantaneous assimilation rates and potential gross photosynthesis of both. Nitrogen available from plant consumption was higher when water was scarce than when it was abundant. Additionally nitrogen uptake was not limited by soil water content. Understanding the reasons of these dynamics would require a more detailed analysis of the processes that belong to the nitrogen and water cycles. A preliminary exploration of these processes shows that the amount of nitrogen lost through leaching is higher when water is abundant. In the nitrogen

unlimited scenarios the plant demand for nitrogen was fulfilled in its totality by entering extra amounts of nitrogen into the system. In these scenarios, the availability of nitrogen allowed the plants to grow more than previously. However, this growth was limited by the water deficit resulting by the increasing demand of water to evapo-transpire.

Figure 4 a and b show partial flow diagrams that depict how the interactions between processes change in response to LAI perturbations for the two scenarios where nitrogen was limited. The different trends observed in these two graphs originated from the water availability in the soil and its effects in stomata conductance. The analysis shows that when water was limited the conductance to CO2 was low and more prompt to changes in LAI than when water was unlimited, producing higher variations in assimilation rate and potential gross photosynthesis. This pattern of variation was also followed by the actual gross photosynthesis, the amount of carbon and nitrogen allocated to the leaves and the total maintenance respiration. On the other hand, when water was abundant the conductance for the sun part of the canopy was always at its maximum value and only the conductance of the shaded part was able to vary, resulting in smaller percent of change of assimilation rate, potential gross photosynthesis actual gross photosynthesis, the amount of carbon and nitrogen allocated to the leaves and the total maintenance respiration.

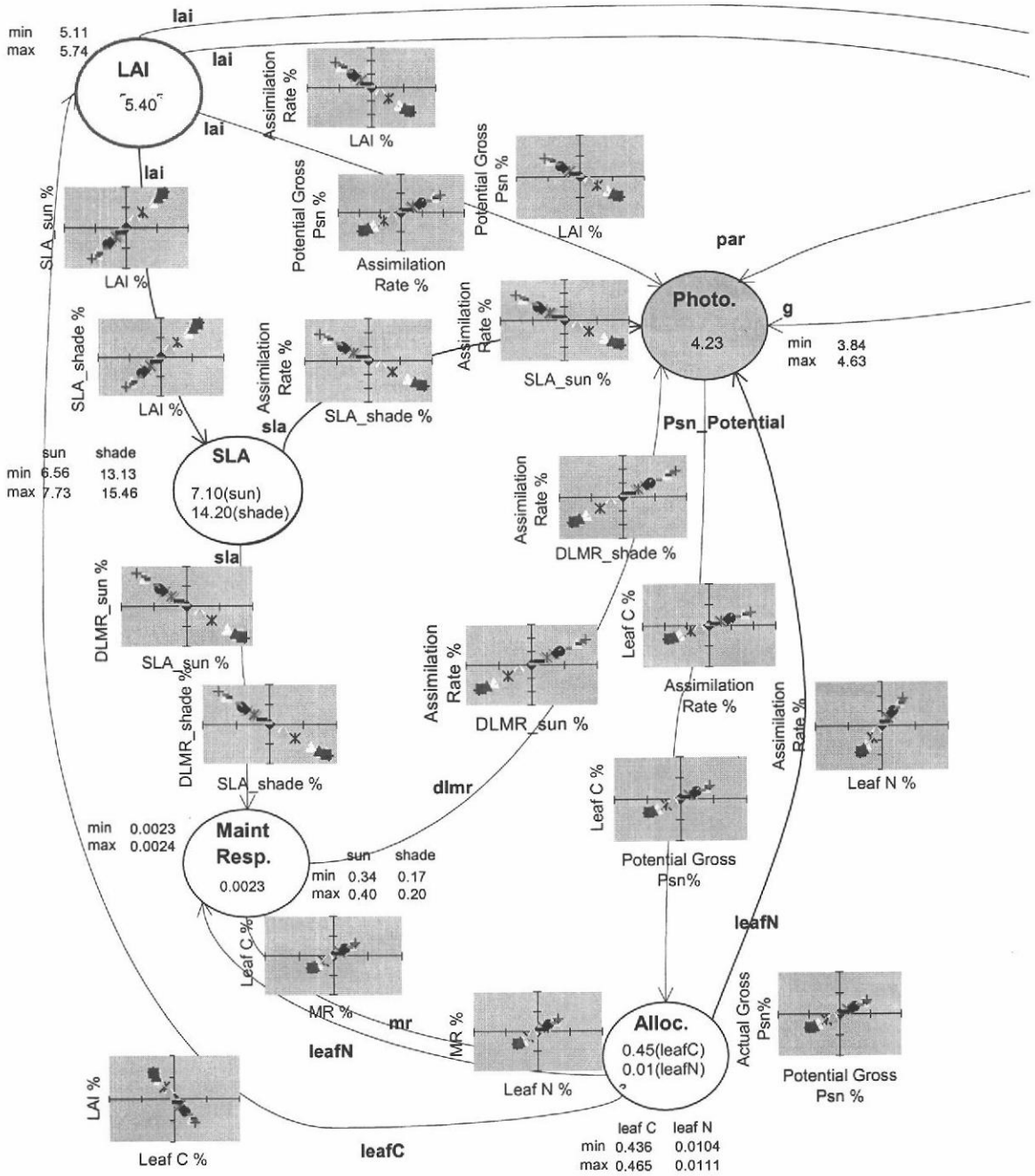


Figure 4.a. Results of perturbing the LAI process in a scenario where water and nitrogen are limited.

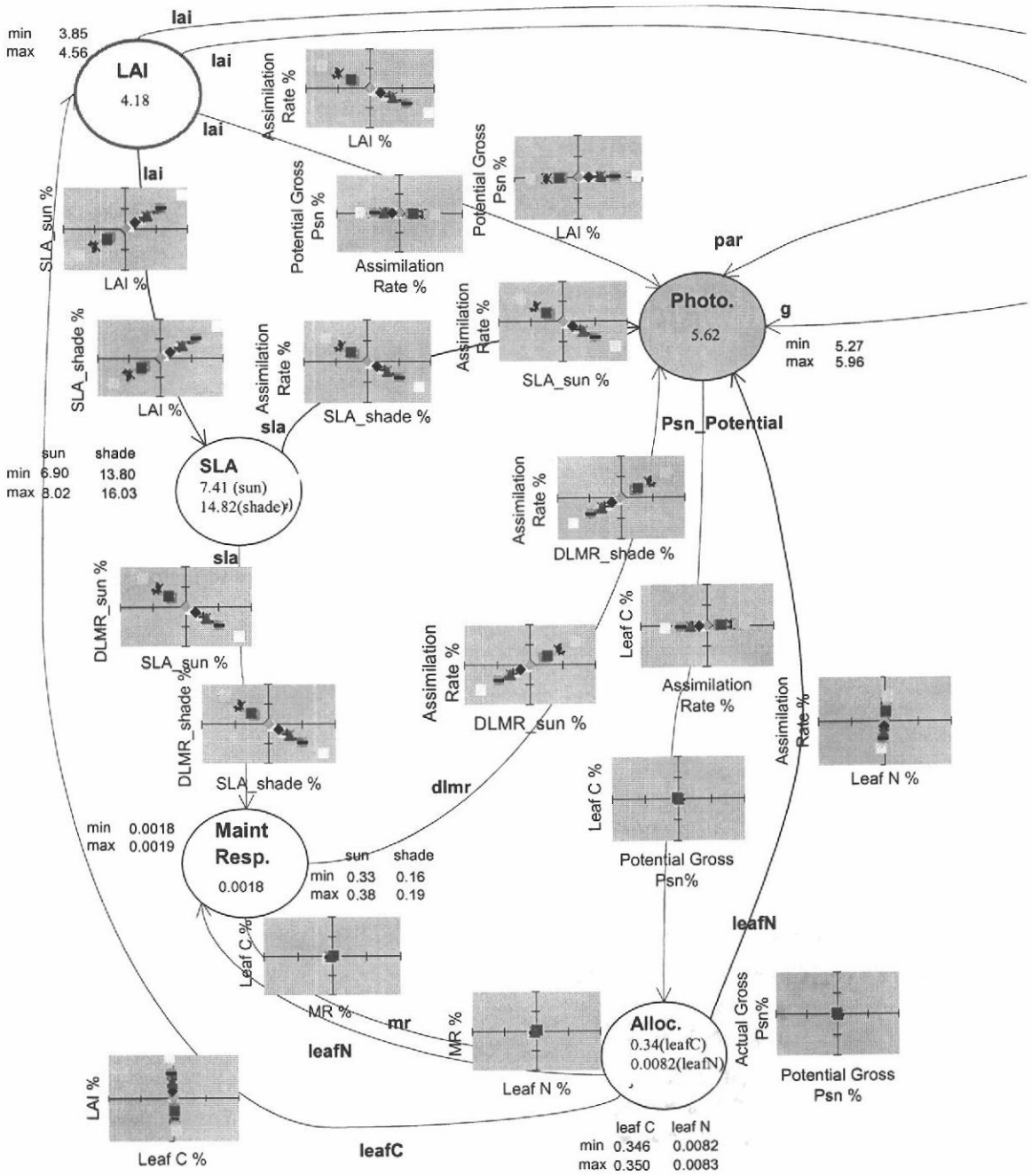


Figure 4.b. Results of perturbing the LAI process in a scenario where water is abundant and nitrogen limited.

NPP values were also affected by variations in LAI. Following a similar pattern to actual gross photosynthesis, these two values diminished as LAI increased. When nitrogen availability was a limiting factor, NPP was the highest when water was limited, with values that varied from 562.33 to 601.76 kgC/m^2. On the other hand, when water was abundant NPP varied from 445.81 to 451.02 kgC/m^2. When nitrogen was unlimited the highest values of NPP were reached in the case where there were no water limitations, with NPP values ranging from 1377.33 to 1444.35 kgC/m^2. In the case were there were water limitations values varied from 678.84 to 767.92 kgC/m^2. Table 7 summarizes how the outputs vary in response to perturbations in the LAI calculation process in all four scenarios.

Variation in output data generated by model perturbations.

The assimilation rate was most sensitive to changes in LAI in the scenarios where water was limited. Here sensitivity values ranged between 2.3 and 2.6 for the cases where nitrogen was unlimited. This means that a unit of change in LAI generated between 2.3 and 2.6 units of change in the assimilation rates. The other scenarios showed a weaker dependency on variations in LAI, with the lower sensitivity values ranging between 0.8 and 0.9 and happening in the scenario where water was abundant and nitrogen limited. The potential gross photosynthesis highest sensitivity values were also observed in the two scenarios where water was limited, with values ranging between 1 and 2 in the case of nitrogen unlimited and 0.8 and 0.9 in the case of nitrogen limited. The scenario where water was abundant and nitrogen limited expressed no sensitivity, meaning that variations in LAI did not affect the outcome of potential canopy photosynthesis.

The actual gross photosynthesis was also most sensitive in the scenarios where water was limited (see Table 5). In the case of nitrogen limitations the sensitivity reached a value that ranged between 0.5 and 0.6; while when nitrogen was unlimited its value ranged from 1.4 to 1.75. Sensitivities when water was abundant were much smaller, reaching a value close to zero when water was abundant and nitrogen limited. NPP was most sensitive when water was limited, with sensitivity values ranging from 0.5 to 1.7. Table 6 shows how NPP responded to LAI variations.

When perturbing LAI, the sensitivity values showed to vary linearly with respect to changes in the LAI process (see Tables 3 to 6) .In most cases the sensitivity values were constant through all the runs, meaning that changes in the LAI process always produced a proportional change in the output analyzed. In another cases, the sensitivity values either diminished or increased in response to changes in LAI. For example, in the scenario where water and nitrogen were limited, the sensitivity values of assimilation rate decreased as the percentage of change in LAI increased. This means that the values of assimilation rates expressed more change when LAI decreased than when it increased. On the other hand, the opposite response is observed in the scenarios where water was abundant and nitrogen unlimited. However, this trends where very small to draw strong conclusions about model behavior in different scenarios.

Table 3. This table shows how the instantaneous assimilation rate responded to variations in LAI under different boundary conditions of nitrogen and water availability. For each scenario three graphs are depicted: (1) the instantaneous assimilation rate absolute values that result from varying LAI, (2) the percent change in instantaneous assimilation rate that result from varying LAI and (3) the sensitivity measure. In all the graphs, the x axis represents the percent change in LAI. The y axis represents the absolute value in the first graph, the percent change in the second graph and the sensitivity value in the third graph as indicated by column 2. The percent change is computed as indicated below. The sensitivity is computed as indicated in equation 1.

$$\textit{Percentage change} = [(O_p - O_b)/O_b]$$

where,

O_b response variable that is subject to analysis
O_p perturbed response variable

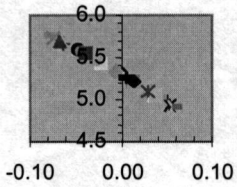
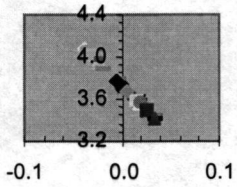
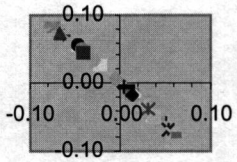
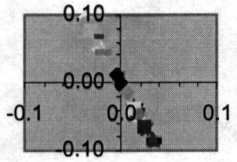
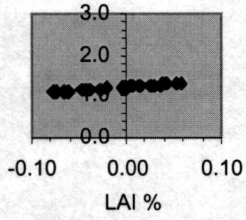
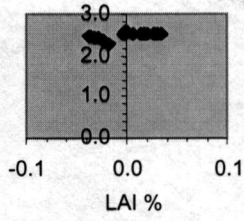
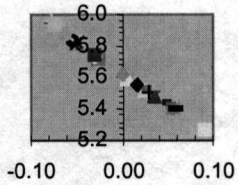
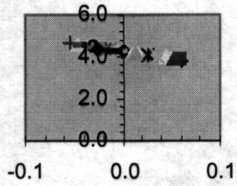
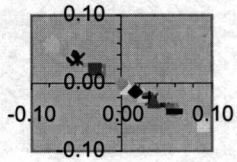
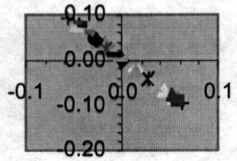
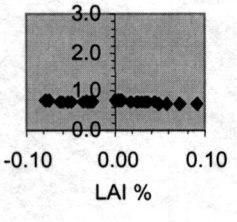
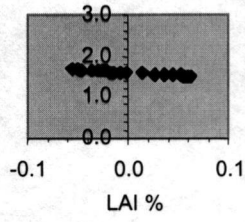
Assimilation Rate in LAI		Water abundant	Water limited
Nitrogen unlimited	**Absolute value**		
	Percentage change		
	Sensitivity		
Nitrogen limited	**Absolute value**		
	Percentage change		
	Sensitivity		

Table 4. This table shows how the potential gross photosynthesis responded to variations in LAI under different boundary conditions of nitrogen and water availability. For each scenario three graphs are depicted: (1) the potential gross photosynthesis absolute values that result from varying LAI, (2) the percent change in potential gross photosynthesis that result from varying LAI and (3) the sensitivity measure. In all the graphs, the x axis represents the percent change in LAI. The y axis represents the absolute value in the first graph, the percent change in the second graph and the sensitivity value in the third graph as indicated by column 2. The percent change and sensitivity values are computed as indicated in Table 3.

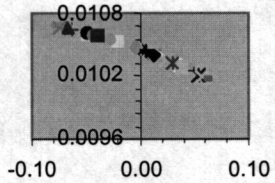
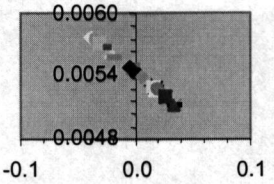
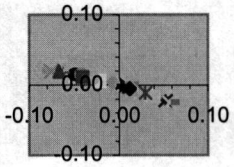
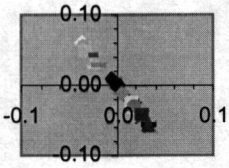
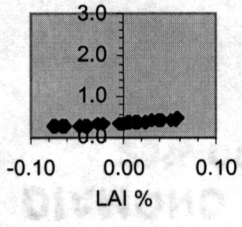
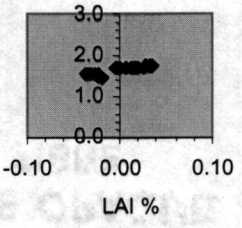
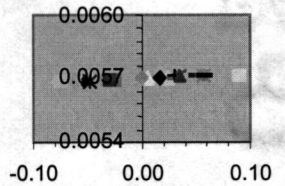
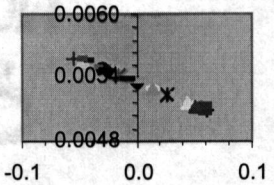
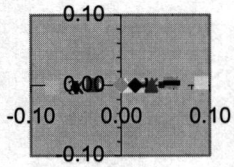
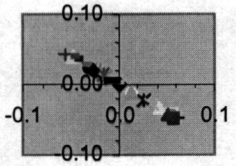
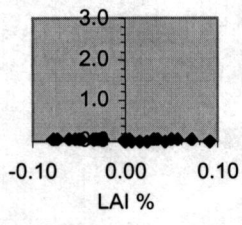
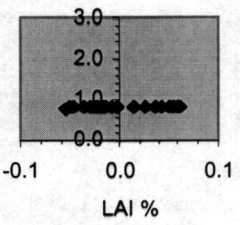
Potential Gross Psn in LAI		Water abundant	Water limited
Nitrogen unlimited	Absolute value		
	Percentage change		
	Sensitivity		
Nitrogen limited	Absolute value		
	Percentage change		
	Sensitivity		

Table 5. This table show how the actual gross photosynthesis responded to variations in LAI under different boundary conditions of nitrogen and water availability. For each scenario three graphs are depicted: (1) the actual gross photosynthesis absolute values that result from varying LAI, (2) the percent change in actual gross photosynthesis that result from varying LAI and (3) the sensitivity measure. In all the graphs, the x axis represents the percent change in LAI. The y axis represents the absolute value in the first graph, the percent change in the second graph and the sensitivity value in the third graph as indicated by column 2. The percent change and sensitivity values are computed as indicated in Table 3.

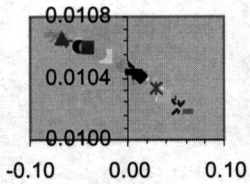
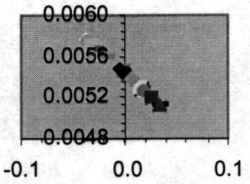
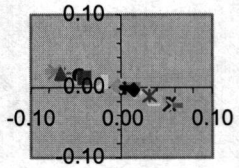
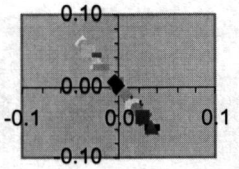
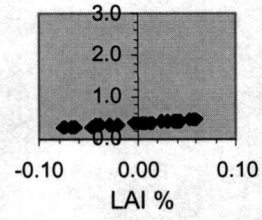
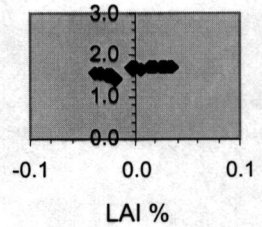
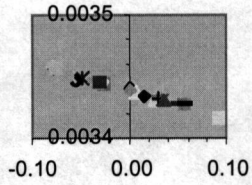
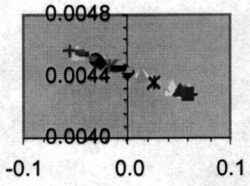
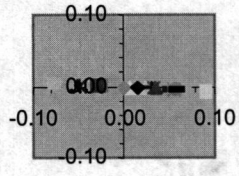
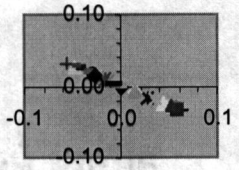
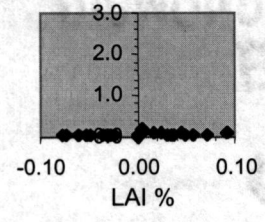
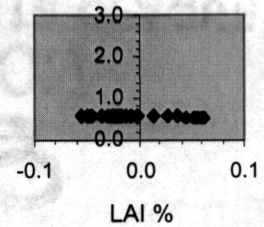
Actual Gross Psn in LAI		Water abundant	Water limited
Nitrogen unlimited	Absolute value		
	Percentage change		
	Sensitivity		
Nitrogen limited	Absolute value		
	Percentage change		
	Sensitivity		

Table 6. This table shows how the NPP responded to variations in LAI under different boundary conditions of nitrogen and water availability. For each scenario three graphs are depicted: (1) the NPP absolute values that result from varying LAI, (2) the percent change in NPP that result from varying LAI and (3) the sensitivity measure. In all the graphs, the x axis represents the percent change in LAI. The y axis represents the absolute value in the first graph, the percent change in the second graph and the sensitivity value in the third graph as indicated by column 2. The percent change and sensitivity values are computed as indicated in Table 3.

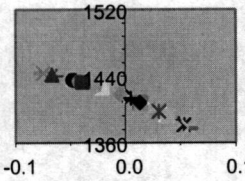
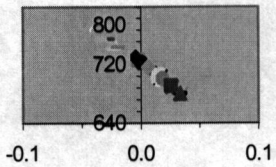
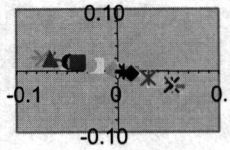
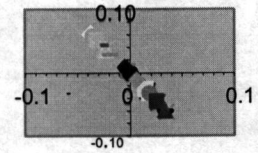
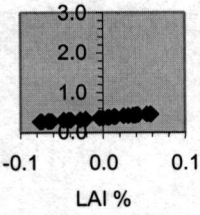
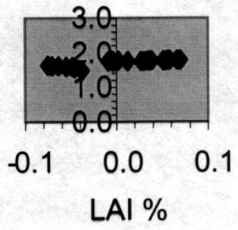
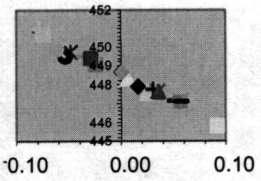
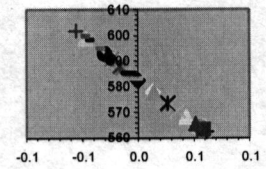
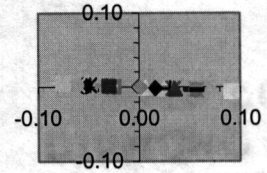
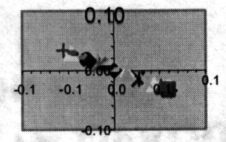
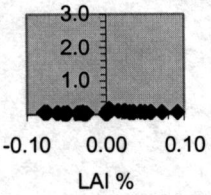
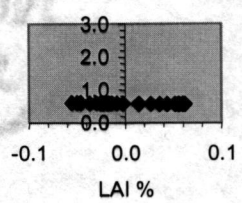
NPP in LAI		Water abundant	Water limited
Nitrogen unlimited	Absolute value		
	Percentage change		
	Sensitivity		
Nitrogen limited	Absolute value		
	Percentage change		
	Sensitivity		

Table 7. These tables show the range of output variations in response to perturbations in the LAI calculation process for four different scenarios. The table's first row and column indicates the scenarios. The second column identifies the output which values are shown. In the first table, columns third and fourth and second and third row, show the baseline, the maximum and the minimum values of LAI for each scenario. In all other tables they show the baseline value and the value of corresponding output when LAI is minimum and maximum.

LAI perturbations		Water abundant		Water limited
Nitrogen unlimited	**LAI**	Baseline	12.42	6.57
		Min	11.47	6.32
		Max	13.15	6.80
Nitrogen limited	**LAI**	Baseline	4.18	5.4
		Min	3.85	5.10
		Max	4.56	5.74

LAI perturbations		Water abundant		Water limited
Nitrogen unlimited	Assimilation Rate (umol/m2/sec)	Baseline	5.30	3.72
		When LAI min	5.75	4.07
		When LAI max	4.9	3.40
Nitrogen limited	Assimilation Rate (umol/m2/sec)	Baseline	5.62	4.239
		When LAI min	5.96	4.634
		When LAI max	5.27	3.884

LAI perturbations		Water abundant		Water limited
Nitrogen unlimited	Potential PSN (kgC/m2/d)	Baseline	0.0104	0.00544
		When LAI min	0.0106	0.00576
		When LAI max	0.0101	0.00511
Nitrogen limited	Potential PSN (kgC/m2/d)	Baseline	0.0057	0.0053
		When LAI min	0.0056	0.0056
		When LAI max	0.00571	0.0051

LAI perturbations		Water abundant		Water limited
Nitrogen unlimited	Actual PSN (kgC/m2/d)	Baseline	0.0104	0.00544
		When LAI min	0.0106	0.00576
		When LAI max	0.0101	0.00511
Nitrogen limited	Actual PSN (kgC/m2/d)	Baseline	0.00344	0.0044
		When LAI min	0.00345	0.0046
		When LAI max	0.00341	0.0043

LAI perturbations		Water abundant		Water limited
Nitrogen unlimited	GPP (kgC/m2)	Baseline	3818.57	1986.52
		When LAI min	3897.92	2104.23
		When LAI max	715.85	1866.90
Nitrogen limited	GPP (kgC/m2)	Baseline	1255.45	1616.39
		When LAI min	1262.34	1668.56
		When LAI max	1247.04	1562.33

LAI perturbations		Water abundant		Water limited
Nitrogen unlimited	NPP (kgC/m2)	Baseline	1415.22	723.76
		When LAI min	1444.35	767.92
		When LAI max	1377.33	678.84
Nitrogen limited	NPP (kgC/m2)	Baseline	448.67	582.32
		When LAI min	451.02	601.76
		When LAI max	445.81	562.33

Maintenance respiration process effects

Maintenance respiration constitutes the basal rate of metabolism. In this model it is calculated as a function of tissue nitrogen concentration, using the same base respiration rate for all live plant compartments and making corrections for temperature. For the leaves, the maintenance respiration is calculated separately for the day and night. The maintenance respiration is used by the *allocation* process to compute the daily net photosynthesis (i.e. by subtracting the carbon lost in respiration from the daily gross photosynthesis). The model also calculates the DLMR_sun and DLMR_shade. This rate is used by the *photosynthesis* process for the calculation of the instantaneous assimilation rate and the daily gross photosynthesis.

The perturbations imparted to the *maintenance respiration* process either increment or reduce the mass of carbon respired by the plant. These perturbations were the same in all four scenarios, however, similarly to what was done when perturbing the *LAI calculation* process, the model responded differently to these changes depending on the nitrogen and water availability of the site. The variations in

maintenance respiration were in the range of +/-14 % when nitrogen was limited and in the range of +/-10 % otherwise.

Alteration in process-process interactions

Variations in maintenance respiration directly affected the *photosynthesis* process. In the scenarios where nitrogen was unlimited, increasing the mass of carbon respired increased the instantaneous assimilation rate, while the opposite happened when nitrogen was a limited resource (see Table 8). When nitrogen was unlimited, the assimilation rate responded to increases in conductance to CO_2 and ppfd. As the maintenance respiration increased, the plant leaf carbon and the LAI diminished, allowing the plant to further open the stomata to transpire, so favoring the intake of CO_2. The variations observed in leaf conductance to CO_2 were on the order of -7% to +12% for the shade leaves and on the order of +/- 6% for the sun leaves when water was limited and +/- 3% when it was abundant.

Reductions in LAI also affected the *radiation uptake* process, decreasing the aPar and increasing the ppfd. The variations in ppfd for the shade part of the leaves were in the order of -7% to 13% when water was abundant and -3% and 9% when water was limited. The ppfd for the sunlit part of the leaves expressed no change, because the sunlit area of leaves always reached its maximum value of 1.

In the two scenarios where nitrogen was limited, the lack of nitrogen played an essential role in constraining the system. Here, the instantaneous assimilation rate responded to increases in leaf nitrogen content. It was when leaf nitrogen content

reached the lowest values that the assimilation rate responded to increases in conductance to CO_2 and ppfd. Yet, the changes in assimilation rate were less than 1% of the baseline value. Table 8 shows how this rate varies through the four different scenarios.

The potential gross photosynthesis was also affected by variations in the process of *maintenance respiration*. This potential was estimated by multiplying the assimilation rate by the day length in the two fractions of the canopy and adding the daytime leaf maintenance respiration. Table 9 shows how the daily gross photosynthesis varies through the four different scenarios.

The potential gross photosynthesis was used to calculate the plant nitrogen demand and based on how much of this demand could be fulfilled the model estimated the actual gross and net photosynthesis. When nitrogen was an unlimited resource, the actual gross photosynthesis was equivalent to those of the potential (see Table 10). However when nitrogen was limited, actual values were smaller than the potential. This reflected the lack of nitrogen available for plant uptake, so the plants had to adjust productivity to the availability of mineral nitrogen. In the two scenarios where nitrogen was limited, as maintenance respiration increased the actual gross photosynthesis also increased while the NPP diminished. The variations observed in the actual gross photosynthesis were in the order of +/-10 %.

GPP followed the same patterns of variation as daily gross photosynthesis did, with equivalent percent change and sensitivity values. GPP was the highest in the scenarios where nitrogen was unlimited, with values that ranged from 3683 to 3839

kgC/m^2 when water was abundant and from 1986 to 1993 kgC/m^2 when water was scarce. The smaller values happened in the scenario where nitrogen was limited and water abundant, with values ranging from 1183 to 1350 kgC/m^2.

NPP showed the biggest variability in the scenarios where nitrogen was unlimited, with values that ranged in +/- 10 % of the baseline value (see Table 11). In the scenarios where nitrogen was limiting, NPP only showed some variability when the respiration reached the highest values. Yet, the percentage of change in the resulting NPP was less than 2%. Similarly to what happened with GPP, NPP was the highest in the scenarios where nitrogen was unlimited, with values that ranged from 1247 to 1554 kgC/m^2, when water was abundant, and from 660 to 770 kgC/m^2, when water was scarce. The smaller values happened in the scenario where nitrogen and water was limited, with values ranging from 446 to 449 kgC/m^2. Table 12 shows the range of output variations resulting from perturbing the *maintenance respiration* process.

Variation in output data generated by model perturbations.

It was under nitrogen unlimited conditions that the instantaneous assimilation rate expressed the highest variability in response to changes in maintenance respiration. When water was limited the sensitivity values ranged from 0.4 to 0.7. This means that a normalized unit of change in maintenance respiration generated a normalized 0.4 unit of change in assimilation rate. When water was abundant the sensitivity values were less than 0.6. On the other hand, in the two scenarios where nitrogen was limited the sensitivity values were smaller than 0.1, indicating that changes in

maintenance respiration had a small repercussion on the values of instantaneous assimilation rate.

Daily gross photosynthesis was most sensitive to increases in maintenance respiration when nitrogen was unlimited and water abundant, with sensitivities exponentially reaching a value of one as respiration increased. The actual gross photosynthesis expressed the same sensitivity when nitrogen was unlimited. However, when nitrogen was a limiting resource, the sensitivities were equal to 0.5. These results highlighted the importance of nitrogen availability in modifying the effects changes in maintenance respiration have in the photosynthesis and allocation of carbon to the plant.

NPP showed to be most sensitive to variations in maintenance respiration in the scenarios where nitrogen was unlimited with values reaching a value of 3 when water was abundant and a value of 2 when water was limited (see Table 11). In the scenarios where nitrogen was limiting, NPP showed to be sensitive only when the respiration reached the highest values. Yet, the sensitivity values were smaller than 0.12.

In the scenarios where nitrogen was an unlimited factor the sensitivity values showed a non-linear trend with respect to changes in maintenance respiration. For example, when nitrogen and water were unlimited the sensitivity values of potential and actual gross photosynthesis and NPP showed an exponential trend. This means that the effects that perturbing maintenance respiration has on these variables depend on the magnitude of the change, being increasingly more important as the maintenance

respiration increases (Tables 9, 10 and 11). However, when nitrogen was unlimited and water limited, sensitivity values did not show the well defined exponential trend mentioned above. Instead they showed a non-linear pattern that revealed how small variations in the maintenance respiration resulted in high sensitivity values, affecting the resulting NPP much more than bigger variations. In the scenarios where nitrogen was limited, sensitivity values either had a value very close to zero or they followed a linear trend, showing that changes in maintenance respiration did not significantly affected the outputs analyzed. One possible hypothesis to explain the non-linear trend in sensitivity values is that they occur when conductance to CO2 and ppfd dominate the photosynthesis; while when the photosynthesis process is dominated by changes in leaf nitrogen concentration this trend is linear. Another possible hypothesis is that there is a Q10 effect (temperature) showing up under nitrogen unlimited conditions but being muted under nitrogen limited conditions; that is temperature dominates in one case and nitrogen in the other.

Table 8. This table shows how the assimilation rate responded to variations in maintenance respiration under different boundary conditions of nitrogen and water availability. For each scenario three graphs are depicted: (1) the assimilation rate absolute values that result from varying maintenance respiration, (2) the percent change in assimilation rate that result from varying maintenance respiration and (3) the sensitivity measure. In all the graphs, the x axis indicates the percent change in maintenance respiration. The y axis represents the absolute value in the first graph, the percent change in the second graph and the sensitivity value in the third graph as indicated by column 2. The percent change and sensitivity values are computed as indicated in Table 3.

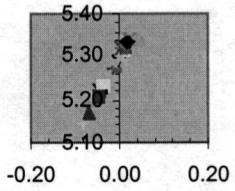
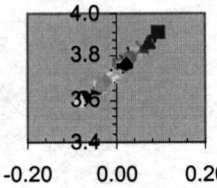
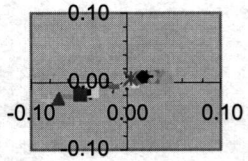
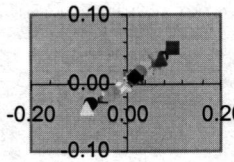
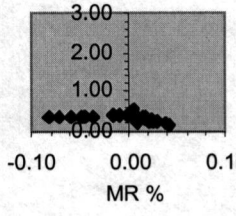
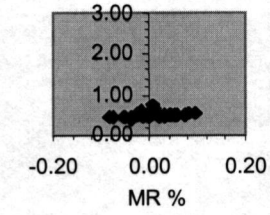
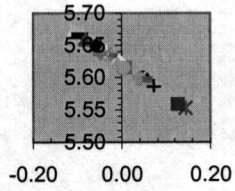
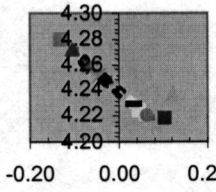
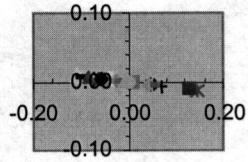
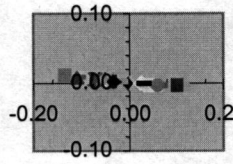
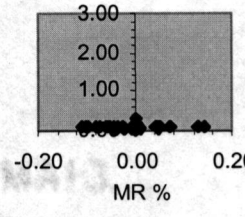
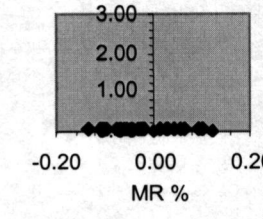
Assimilation Rate in MR		Water abundant	Water limited
Nitrogen unlimited	Absolute value		
	Percentage change		
	Sensitivity		
Nitrogen limited	Absolute value		
	Percentage change		
	Sensitivity		

Table 9. This table shows how the potential gross photosynthesis responded to variations in maintenance respiration under different boundary conditions of nitrogen and water availability. For each scenario three graphs are depicted: (1) the potential gross photosynthesis absolute values that result from varying maintenance respiration, (2) the percent change in potential gross photosynthesis that result from varying maintenance respiration and (3) the sensitivity measure. In all the graphs, the x axis indicates the percent change in maintenance respiration. The y axis represents the absolute value in the first graph, the percent change in the second graph and the sensitivity value in the third graph as indicated by column 2. The percent change and sensitivity values are computed as indicated in Table 3.

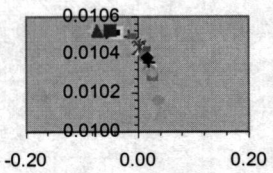
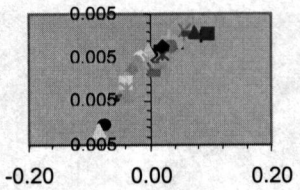
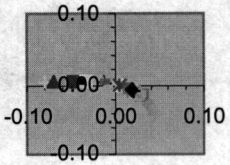
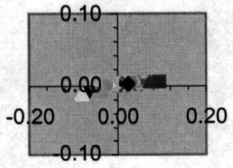
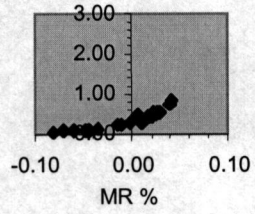
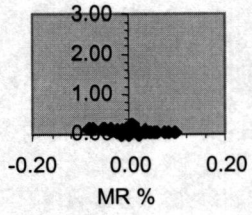
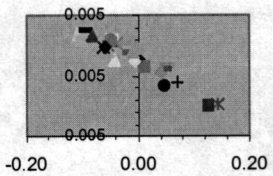
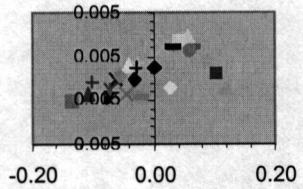
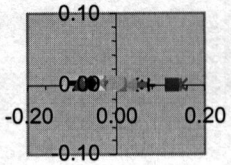
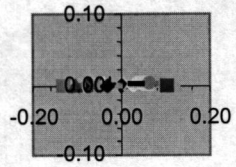
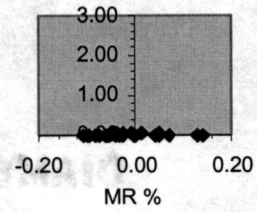
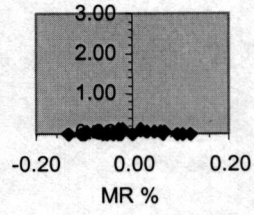
Potential Gross Psn in MR		Water abundant	Water limited
Nitrogen unlimited	Absolute value		
	Percentage change		
	Sensitivity		
Nitrogen limited	Absolute value		
	Percentage change		
	Sensitivity		

Table 10. This table shows how the actual gross photosynthesis responded to variations in maintenance respiration under different boundary conditions of nitrogen and water availability. For each scenario three graphs are depicted: (1) the actual gross photosynthesis absolute values that result from varying maintenance respiration, (2) the percent change in actual gross photosynthesis that result from varying maintenance respiration and (3) the sensitivity measure. In all the graphs, the x axis indicates the percent change in maintenance respiration. The y axis represents the absolute value in the first graph, the percent change in the second graph and the sensitivity value in the third graph as indicated by column 2. The percent change and sensitivity values are computed as indicated in Table 3.

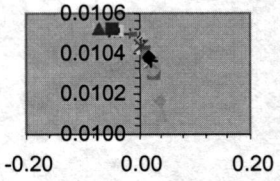
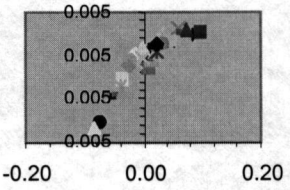
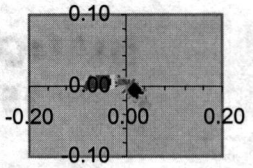
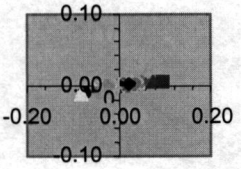
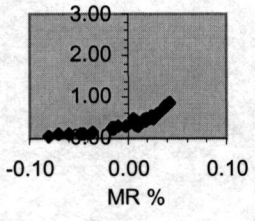
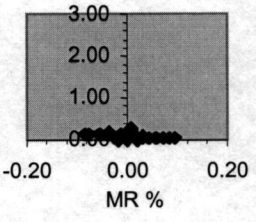
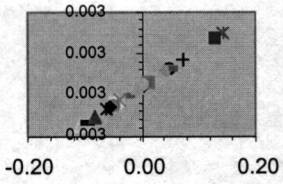
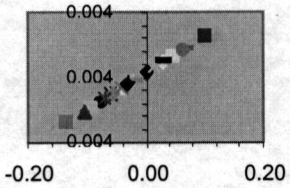
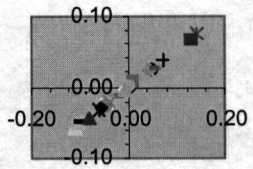
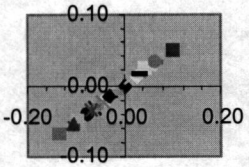
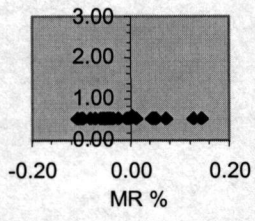
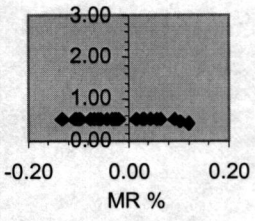
Actual Gross Psn in MR		Water abundant	Water limited
Nitrogen unlimited	Absolute value		
	Percentage change		
	Sensitivity		
Nitrogen limited	Absolute value		
	Percentage change		
	Sensitivity		

Table 11. This table shows how the NPP responded to variations in maintenance respiration under different boundary conditions of nitrogen and water availability. For each scenario three graphs are depicted: (1) the NPP absolute values that result from varying maintenance respiration, (2) the percent change in NPP that result from varying maintenance respiration and (3) the sensitivity measure. In all the graphs, the x axis indicates the percent change in maintenance respiration. The y axis represents the absolute value in the first graph, the percent change in the second graph and the sensitivity value in the third graph as indicated by column 2. The percent change and sensitivity values are computed as indicated in Table 3.

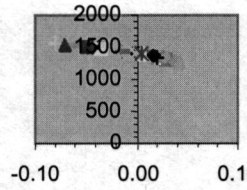
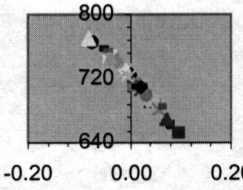
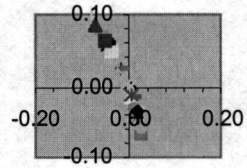
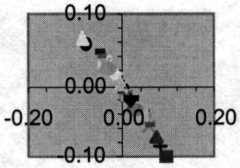
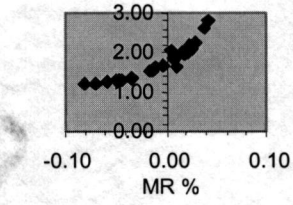
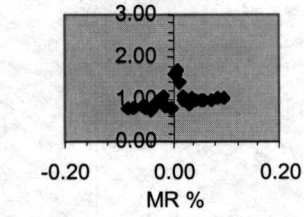
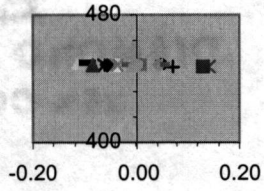
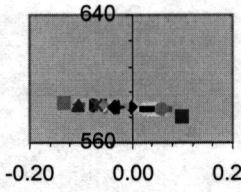
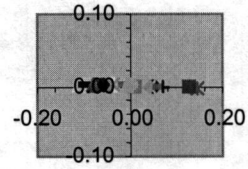
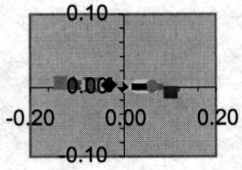
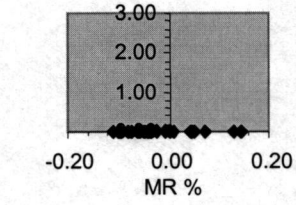
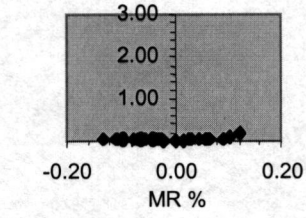
NPP in MR		Water abundant	Water limited
Nitrogen unlimited	Absolute value		
	Percentage change		
	Sensitivity		
Nitrogen limited	Absolute value		
	Percentage change		
	Sensitivity		

Table 12. These tables show the range of output variations in response to perturbations in the Maintenance Respiration calculation process for four different scenarios. The table's first row and column indicates the scenarios. The second column identifies the output which values are shown. Columns third and fourth and second and third row, show the baseline, the maximum and the minimum values for each scenario.

MR perturbations		Water abundant	Water limited
Nitrogen unlimited	LAI	Baseline 12.4 Min 10.1 Max 13.6	6.57 6.05 6.97
Nitrogen limited	LAI	Baseline 4.18 Min 4.16 Max 4.19	5.4 5.28 5.44

MR perturbations		Water abundant	Water limited
Nitrogen unlimited	Photo. Rate (umol/m2/sec)	Baseline 5.30 When MR min 5.15 When MR max 5.35	3.72 3.6 3.9
Nitrogen limited	Photo. Rate (umol/m2/sec)	Baseline 5.62 When MR min 5.55 When MR max 5.62	4.239 4.23 4.28

MR perturbations		Water abundant	Water limited
Nitrogen unlimited	Potential PSN (kgC/m2/d)	Baseline 0.0104 When MR min 0.0105 When MR max 0.010	0.00544 0.0053 0.0054
Nitrogen limited	Potential PSN (kgC/m2/d)	Baseline 0.0057 When MR min 0.00569 When MR max 0.0057	0.0053 0.0053 0.0053

MR perturbations		Water abundant	Water limited
Nitrogen limited	Actual PSN (kgC/m2/d)	Baseline 0.0034 When MR min 0.0032 When MR max 0.0037	0.00544 0.0053 0.0054
Nitrogen limited	Actual PSN (kgC/m2/d)	Baseline 0.00344 When MR min 0.0032 When MR max 0.0037	0.0044 0.0041 0.0046

MR perturbations		Water abundant	Water limited
Nitrogen limited	GPP (kgC/m2)	Baseline 3818.57 When MR min 3683.7 When MR max 3839.8	1986.52 1960.8 1993.1
Nitrogen limited	GPP (kgC/m2)	Baseline 1255.45 When MR min 1183.3 When MR max 1350.2	1616.39 1504.6 1701.8

MR perturbations		Water abundant		Water limited
Nitrogen unlimited	NPP (kgC/m2)	Baseline When MR min When MR max	1415.22 1247.0 1554.1	723.76 660.9 770.7
Nitrogen limited	NPP (kgC/m2)	Baseline When MR min When MR max	448.67 446.4 449.5	582.32 584.7 568

Discussion

The increasing complexity of models and the rapidity with which they evolve has put demands on the capacity to effect sensitivity analyses. Such increased complexity entails more than increased c*omplicatedness* (sensu Allen, 1999). True, improved computational capabilities have allowed modeler's to build and test more intricate models simulating ecosystem behavior, however, there has been also a significant shift in the perception of how ecosystems function. Biophysical systems are now being modeled as complex systems which are not so much complicated as they are sensitively tuned networks where the relationships between elements (e.g., initial conditions, processes, and process interactions and rates) are key to understanding system behavior. For this reason, there is increasing need to develop sensitivity analyses which parallel model development and complexity. Simulated processes, like the empirical information used to tailor input or parameter values, are subject to many sources of uncertainty. The sensitivity analysis methodology presented recognizes a process (i.e., a functional unit that simulates a specific ecosystem action), has its own associated ecological meaning and relevance different than its constituent parts. As such, evaluation of model behavior at the process level, is incomplete when rendered solely at the level of parameters or input variables;

process behavior is not always be captured by analyses executed at the parameter or input levels.

It is important to make the distinction between uncertainty assessment targeted at the parameter level versus uncertainty assessment aimed at the process level *which can utilize parameter perturbations to evaluate model behavior change.* In the case study described here, this distinction was illustrated. While parameters were used as a vehicle to perturb the system, they were not subject to any analysis themselves. The process level analysis focuses on how the process behavior is affected and is not concerned with the relationship between parameter and model behavior. When perturbing the LAI process the perturbations were carried on through the parameter that represented the average projected leaf area. These perturbations generated either an increase or decrease in the resulting LAI. The sensitivity analysis focused on understanding how these variations in LAI, rather than variations in the average projected leaf area parameter, affected model behavior. This was carried out by analyzing the repercussions the increase or decrease in LAI had on the functioning of other processes as well as on the final outcome of the model. The information gathered explained model behavior as triggered by the *LAI calculation* process perturbations providing the modeler a detailed description of the internal workings of the carbon cycle. This enables the modeler to make inference on the appropriateness of function representation itself.

Subsequently, the process-level approach shifts the modeler's attention from result precision to an evaluation of the question underlying the modeling exercise itself. This approach works at the process level to explicitly inform about how processes

interact to generate a particular output. With this information the modeler can evaluate if such behavior is the one intended to be modeled and s/he can judge the resemblance of the model with the real ecosystem (e.g., is the evapo-transpiration increasing as LAI increase? how do changes in *LAI calculation* process affect the *photosynthesis* process? Is nitrogen constraining plant growth as LAI increases? Are the resulting values of LAI realistic?). The modeler can also evaluate whether or not the output sensitivity is the expected one (e.g., NPP values were more sensitive to LAI process when water was a limiting factor) and can determine the contribution the formulation of the LAI calculation process has in model output variability.

In the case study, LAI values were impacted differently in all four scenarios, producing different model dynamics. When varying LAI, the photosynthesis was most sensitive when water was a limiting factor, revealing the dependency the predictions from photosynthesis had on the water cycle. It was in those scenarios were the stomata aperture was most limited by water deficit which resulted in the lower assimilation rates and potential photosynthesis. Similarly, the case study results also showed that nitrogen availability also played a fundamental role in this model, by adjusting the actual gross photosynthesis on the basis of nitrogen constraints. Since the cycle of nitrogen was closely linked to that of water, variations in soil water content also affected the availability of nitrogen.

From the two scenarios that were nitrogen limited, the one that was also water limited reached higher values of actual gross photosynthesis and NPP than the one that was water abundant. Thus in this model, productivity was benefited when nitrogen and water were limited and it diminished when the system had more water.

These results were non-intuitive and confronted the modeler with the questions of whether or not this was the desired behavior. These results suggest further investigation of how nitrogen and water cycles are coupled, how the availability of nitrogen is conditioned by soil water content and how nitrogen uptake is limited by water availability. Particularly, the modeler would need to investigate how nitrogen availability can be affected by water through the rates of microbial decomposition or by being carried out of the system through leaching.

In the case of perturbing the maintenance respiration, the analysis also elucidated non-intuitive relationships among the processes. For example, while it was expected that incrementing the maintenance respiration decreases the assimilation rate, this only happened when nitrogen was a limiting factor. When nitrogen was unlimited increasing maintenance respiration was associated with increases in assimilation rate. Additionally, the analysis showed that photosynthesis responded to conductance to CO2 and ppfd in some cases and to nitrogen in others, and it pointed out that these drivers may switch depending on the leaf nitrogen concentration. These results called for further investigation of the nitrogen cycle and allowed to pose hypothesis that could explain this behavior. For example, the modeler may be interested in knowing the conditions under which leaf nitrogen concentration becomes less important for photosynthesis than CO2 and ppfd and s/he can pose hypothesis about this behavior. One possible hypothesis is that this happens when the plant growth is limited by the lack of available nitrogen. This availability is influenced by the inputs (i.e. nitrogen deposition) and outputs (i.e. nitrogen leaching) of nitrogen in the system as well as by microbial immobilization and decomposition. This implies that nitrogen in the system may not be available for plant consumption

due to immobilization, calling for further investigation of the process that models the nitrogen competition between plants and microbes as well as the one that models the decomposition. The modeler could also respond to questions such as, given that the competition for nitrogen uptake seems to be so important, what would happen if it were modeled differently? Would the maintenance respiration perturbations affect the model in the same way? How would the actual behavior change if nitrogen uptake is linked with water uptake? Discerning which is the situation that is more convenient depends on the modeler's scientific capability to identify the best scenario.

Higher order sensitivity analyses such as process-level analyses are not meant as a substitute for traditional parameter-level methods. Ideally, this type of approach expands the context for helping the modeler evaluate the significance of variation and uncertainty. When perturbations are generated by modifying parameters, process based sensitivity analysis provides a complementary tool for effecting parameter sensitivity analysis. To implement the analysis in such a way, the modeler is only required to collect information about changes in output with respect to variation in processes as well as with respect to variations in parameters. At the system level, the methodology provides information about high level interactions that emerge from varying processes; at the parameter level, information concerning the parameter-process relationship becomes the focus of investigation.

Overall, the methodology presented here differs from a traditional sensitivity analysis, and hence responds to the theoretical shift to complex system analysis, in two main ways. First it focuses on evaluating how the model responds to variations in a

process or set of processes, as opposed to how the model responds to variations in parameters, state variables, or input data. Second, the analysis is not only limited to generate information about how variations in a process affect a particular output of the model, but also it aims to understand the repercussions these variations have on the internal workings of the model as characterized by process interactions. In application and discussion of this methodology, a model is conceived as a mechanism that is composed multiple interacting processes. As a result of these interactions, a system of dynamic phenomena, not explicitly encoded in the model, is generated. Thus, analyzing model behavior becomes a task that not only involves the evaluation and understanding of output variability but also of *how* the output is generated; overall expressing a concern for the relationship that exists between process interactions and the emergent properties of the model.

While we focused on intra-model dynamics, the same approach applies to an even higher level of organization, namely inter-model comparisons where it is possible to evaluate not only model the uncertainties embedded within a given model but the uncertainty associated in the process of conceptualization by the modeler him/herself. This higher-order examination of sources and nature of uncertainty provides a way to examine fundamental assumptions upon which various models are built. It opens to dialogue to explore different conceptualizations and the very objectives that drive the modeling efforts. This is particularly important when dealing with complex systems where changes in local conditions can have a big impact in global conditions and where it is incomplete not to consider sources and nature of uncertainty at various scales. Such inter-model comparisons as Vegetation-Ecosystem Modeling and Analysis Project (VEMAP) would be likely candidates for

this approach. The VEMAP project aims to understand ecosystem dynamics by comparing the results of various global vegetation models (VEMAP1). This project was conceived under the premise that by running the model under the same input conditions the differences in results are driven by the difference in model assumptions. The application of this methodology would provide the additional advantage of explicitly showing where two models coincide and where they diverge, making the comparison more rigorous.

Conclusions

Process-level sensitivity analyses are designed to address the critical question of a model's sensitivity to assumptions in processes and model structure. It is conceived under the premise that a good modeling practice does not only entail building a model able to reproduce observable data, it also involves building a model able to capture the structure of the real ecosystem and understanding how sensitive a model is to different assumption. The information provided by this methodology can be used to verify that the way in which the model operates is congruent with our understanding of the system being modeled and that it reflects the way in which the real ecosystem operates to produce its behavior.

Acknowledgements

Biome-BGC version 4.1.1 was provided by the Numerical Terradynamic Simulation Group (NTSG) at the University of Montana. NTSG assumes no responsibility for the proper use of Biome-BGC by others. Author: Peter E. Thornton. Source BGC.

Numerical Terradynamic Simulation Group, School of Forestry, University of

Montana Missoula, MT 59812 USA.

References

Allen, T.F.H, J.A. Tainter and T.W. Hoekstra. 1999. Supply-side sustainability: Systems Research and Behavioral Science, 16:404–427.

Beck, B. 2002. Model evaluation and performance. Edited by A.H. El-Shaarawi and W. W. Piegorsch. Encyclopedia of Environmetrics, 3: 275-1279.

Bradshaw, G.A. and J. G. Borchers. 2000. Uncertainty as information: narrowing the science-policy gap. Conservation Ecology, 4(1):7. [online] `URL://www.consecol.org/vol4/iss1/art7

Campolongo F., A. Saltelli, T. Sorensen and S. Tarantola. 2000. Hitchhiker's Guide to Sensitivity Analysis. In Sensitivity Analysis, Edited by A. Saltelli, K. Chan and E.M.Scott. Willey Series in probability and statistics, pp 15-47.

Fishwick, P.A. 1988. The Role of Process Abstraction in Simulation. IEEE Transactions on systems, man and cybernetics, 18(1):18-38.

Funtowicz, S. O. and J. R. Ravetz. 1990. Uncertainty and quality in science for policy. Kluwer Academic, Dordrecht, The Netherlands, 232 pp.

Gardner, R. H., V. H. Dale and R. V. O'Neill. 1990. Error propagation and uncertainty in process modeling. In Process Modeling of Forest Growth Responses to Environmental Stress. Eds. R.K. Dixon, R.S. Meldahl, G.A. Ruark and W.G.Warren. Timber Press, Portland, OR, pp 208-219.

Green, E.J., MacFarlane, D.W. and H.T.Valentine. 2000. Bayesian synthesis for quantifying predictions from process models. Tree Physiology, 20:415-419.

Haefner, J. W. 1996. Modeling Biological Systems, principles and applications. Chapman and Hall, 473 pp.

Hann, D.W.,A.S. Hester and C.L. Olsen. 1997. ORGANON user's manual: Edition 6.0. Department of Forest Resources, Oregon State University, Corvallis, OR, 133 pp.

Kann, A. and J.P. Weyant. 2000. Approaches for performing uncertainty analysis in large scale energy/economic policy models. Environmental Modeling and Assessment. 5(2000) 29-46.

Katz, R. W. 1999. Techniques for estimating uncertainty in Climate Change scenarios and Impact studies. In: Representing Uncertainty in Climate Change Scenarios and Impact Studies. Proceeding of the EU Concerted Action Initiative ECLAT-2 Workshop 1, Helsinki, Finland, April 16th - 18th, 1999, Climatic Research Unit, Norwich, UK., pp 38-53.

Klir, G.J. and M.J. Wierman. 1999. Uncertainty-based information. Elements of a generalized information theory. Second corrected edition. Physica-Verlag Heidelberg, Germany. Pp 169.

Landsberg, J.J. and R.H. Waring. 1997. A generalized model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. Forest Ecology and Management, 95:209-228.

McMurtrie, R.E. 1985. Forest productivity in relation to carbon partitioning and nutrient cycling: A mathematical model. In Attributes of trees as crop plants, Cannell, M.G.R. and J.E.Jackson (eds.) Inst. of Terrest. Ecol. Abbots Ripton, Huntington, U.K.

Oderwald, R. G. and R. P. Hans. 1990. A methodology for constructing, estimating coefficients of, and testing process models. Forest Growth: Process Modelling of Responses to Environmental Stress, pp. 200-207.
Oderwald, R. G. and R. P. Hans. 1993. Corroborating models with model properties. Forest Ecology and Management, 62: 271-283.

Omlin, M. and P. Reichert. 1999. A comparison of techniques for the estimation of model uncertainty. Ecological Modelling, 115:45-59.

Parysow, P., Gertner, G. and Westervelt, J. 2000. Efficient approximation for building error budgets for process based models. Ecological Modelling, 135: 111-125.

Reynolds, J. H. and Ford, E.D. 1999. Multi-criteria assessment of ecological process models. Ecology, 80(2): 538-553.

Saltelli, A. 2000. What is Sensitivity Analysis? In Sensitivity Analysis, Edited by A. Saltelli, K. Chan and E.M.Scott. Willey Series in probability and statistics, pp 3-13.

Thornton, P.E. 1998. Description of a numerical simulation model for predicting the dynamics of energy, water, carbon and nitrogen in a terrestrial ecosystem. Ph.D. dissertation. University of Montana, Missoula, MT 280 pp. [Available from Mansfield Library, University of Montana Missoula, MT 59812].

VEMAP1. Kittel,T.G.F., N.A.Rosenbloom, T.H.Painter, D.S.Schimel, J.M.Melillo, Y.Pan, D.W.Kicklighter, A.D.McGuire, R.P.Neilson, J.Chaney, D.S.Ojima, R.McKeown, W.J.Parton, W.M.Pulliam, I.C.Prentice, A.Haxeltine, S.W.Running, L.L.Pierce, R.R.Nemani, E.R.Hunt, T.M.Smith, B.Rizzo, and F.I.Woodward. 1995. The VEMAP integrated database for modelling United States ecosystem/vegetation sensitivity to climate change. Journal of Biogeography, 22:857-862.

Zimmermann, H.-J. 2000. An application oriented view of modeling uncertainty. European Journal of Operational Research, 122:190-198.

Chapter 4

A Component-Based Approach for the Development of Ecological Simulations

Ecological Modelling
P.O. Box 181
1000 AD Amsterdam
The Netherlands

This paper will be sent for publication to Ecological Modelling

Introduction

Ecological simulations are representations of ecological systems. They reproduce the relevant characteristics of the system and allow us to gain some insights into their behavior in unobserved situations (predictions). Ecological systems are complex and their processes, intricate. Thus, simulation models are based on our understanding of the system, but most definitely they highlight the uncertainties associated with that system.

Ecosystems are complex and its study produces many types of uncertainties of different origins, for these reasons the building of simulations in most cases entails a gradual process that deals with few problems at a time. In our view, the simulation approach to be used in model development needs to be functionally consistent with the process itself. In this paper we suggest that the new programming paradigm promoted by component-based frameworks, has some characteristics that makes it particularly appropriate for the development of ecological simulation models.

Although in principle, building a simulation model may seem a straightforward process, the complexity of ecological systems confronts the modeler with one or more problems in constructing simulation algorithms. One such problem encountered in the conceptualization process is devising ways to cope with system and data intricacy, which arises when there is an abundance of information available that describes the ecosystem to be simulated (Zimmermann, 2000). Model formulation requires the processing of large amounts of information and often the assimilation of diverse theories. One way to address the problem of data intricacy is to reduce complexity to

a manageable level. Reducing complexity implies simplifying the ecosystem description by focusing on those characteristics that are perceived as most important for capturing ecosystem dynamics (Zeigler, 1984; Klir, 1991; Allen and Hoekstra, 1992). Simplifications can be carried out at different levels of detail, where highly abstract specifications can be used to build very simple descriptions that are then refined to generate more complex ecosystem representations. From this perspective, developing an appropriate simulation becomes a careful balance between including too much and too little information.

It has been common practice in ecology to approach complexity by developing a simulation incrementally. Initially, a very simple simulation is constructed, then gradually refined and improved to produce results that reflect a more realistic portrayal of the ecosystem process being modeled. The modeler uses this iterative method to progressively increase their understanding of the modeled system and to determine what changes may be necessary (Baumann, 2000). For example, the modeler may model potential gross photosynthesis (GPP) by estimating GPP as a linear function of aPAR. S/he can gradually add complexity to this model by adding the restrictive constraints of the environment that reduce GPP due to the effects of soil water, low temperature and vapor pressure deficit (Landsberg, 1986). The modeler may also want to compare her/his formulation with another one that treats photosynthesis with more detail as it is the Farquhar model mentioned above.

The other potential difficulty is encountered when there is too little rather than too much data—when there is a lack of information, and sometimes lack of consensus, about the characteristics of the ecological phenomena to be simulated (e.g., how

plants allocate carbon). When there is uncertainty about how a function operates, often there are several competing models of a given ecological process. Under these circumstances, many different descriptions of the system can be formulated, generating more than one possible conceptualization. To address this problem, certain assumptions must be made about the behavior of the ecosystem and its most relevant features. It is also necessary to investigate how competing hypotheses may affect predictions, and how they may limit the applicability of simulations to specific domains. For example, a model that precisely describes the radiation distribution within plant canopies may be difficult to achieve because it requires detailed information about the canopy architecture, the angular distribution of the incident radiation and the spectral properties of the leaves that may not be available. Under this circumstances, the modeler needs to make assumptions about the canopy and it most relevant characteristics (e.g., leaves are horizontal distributed, radiation is uniform through the canopy, etc.). S/he also would need to consider how different assumptions may affect model behavior (e.g., leaves are horizontal distributed versus leaves are distributed in different angles).

Developing a simulation model

The development of an ecological simulation entails three main activities: *conceptualization*, *implementation* and *evaluation*. Conceptualization is the process that generates a description of an ecological process based on ecological theory. In the case of the complex set of interactions and processes that characterize most ecosystems, conceptualization consists of identifying the specific component subsystems needed to describe the ecosystem adequately, formulating a model for

each subsystem, and defining the interactions between these subsystems. Modeled concepts may be fairly simple (e.g., a model that estimates productivity as a linear function of absorbed photosynthetically active radiation, aPAR, as suggested by Jarvis and Leverenz, 1983) or very complex (e.g., a model that estimates productivity considering explicitly how photosynthesis is affected by both the physical and biochemical processes, such as light and dark reactions and the rate at which CO2 in ambient air is supplied, as suggested by Farquhar et al., 1989). Implementation is the construction process that transforms the conceptual model into the actual physical model, the simulation program (Houston and Norris, 2001). This involves writing the computer code and algorithms that render the conceptual model into an executable computer representation. The final step, evaluation, tests the behavior of the simulation model for adequacy and quality.

However, the simulation for many ecological problems would require an iterative process of model formulation, more like a trial and error approach, where modules at different levels of detail are considered in conjunction with different assumptions and hypotheses about how the modeled ecosystem works. As a result, the simple sequence of steps of conceptualization-implementation-evaluation, in reality becomes a series of cycles of conceptualization, implementation, re-conceptualization, code modification, implementation, etc. This iteration typically occurs over the course of time as new information and ideas are generated and subsequently used to modify existing code. In the most extreme cases, key components of the model may be replaced with new modules to reconfigure the model. Large models such as global circulation and biogeochemical models are good examples, where typically a standing model is created, and later modified in pieces as the theory behind their

construction is refined. For example, the biochemical model Forest-BGC (as described in Running and Coughlan, 1988) was first developed to simulate ecosystem processes in forest biomes, then its assumptions were modified to account for processes in both forest and non-forest biomes (Biome-BGC, Thornton, 1998). Presently, Biome-BGC is being further refined into a more complex simulation model (BIOMAP, Neilson in prep) that includes multiple life forms and soil layers, and considers competition for light, water and nitrogen among the life forms.

In short, modeling today is a complex and dynamic activity. Unfortunately, current modeling paradigms (e.g., procedural programming, object-oriented programming) fail to offer good solutions for building simulation models. Among other reasons, these paradigms are not adequate for reproducing the way in which the modeler progresses from observation, to conceptualization, and then to implementation and evaluation. Current modeling paradigms do not provide sufficient means of managing high degrees of complexity. For example, implementation of incremental development is problematic and not well supported. Resulting computer representations are difficult to understand and do not provide the modeler with the flexibility required to add, reduce or change the details of representation to allow a simulation model to evolve into a more complex formulation. Additionally, current modeling paradigms rarely permit easy investigation of the effects that different sets of assumptions have on the prediction of the model. To investigate these effects with current paradigms implies writing code for the different alternatives and bound them when compiling or linking the model, requiring a detailed knowledge of how the new alternative will interact with the rest of the simulation. What we suggest is needed, is a modeling paradigm which mirrors the theory and process of modeling of ecosystems in which ecological

modelers engage; a modeling paradigm that can accommodate to the continuous changes in knowledge and understanding of ecosystems.

Comparison of programming paradigms

There are three main types of programming paradigms: procedural, object-oriented and component-based. Each paradigm supports a different style of programming and imparts different characteristics to the resulting simulation. Table 1 compares the characteristics of these paradigms.

Table 1. Comparison of procedural, object-oriented and component-based frameworks programming. The shaded rows show the characteristics discussed in this paper.

Model Characteristic	**Procedural**	**Object Oriented**	**Component/ Frameworks**
Management of software development (capabilities for dealing with complexity)	Poor support for modeling maintenance, evolution and incremental development.	Improve software development via "encapsulation". Data hidden, protection limit pubic exposure.	Well-defined public interfaces. High level support for common tasks. Easy maintenance.
Representation of ecosystem properties (processes, entities)	Only supports process level descriptions.	Allows tight coupling of process/state representation.	Accepts either procedural or object oriented representations.
Model reusability (plug and play capabilities)	Low	Moderate	High
Analysis of model results (sensitivity analysis, uncertainty analysis, visualization, etc.)	Low	Low/Moderate	High

In a procedural paradigm, ecosystems are described exclusively in terms of processes, which constitute the building blocks of a simulation model. For example, in a vegetation model some of these processes could be: maintenance and growth respiration, carbon allocation, photosynthesis, etc. Each process is an algorithm that performs a particular set of actions, implemented as a function. Data are represented by local or global variables. Under this paradigm a simulation is a collection of functions that consists of sequential calls to the different functions that process arguments and return a value. An important limitation of this paradigm is that data and functions that are logically grouped together are physically separated in the code. This separation makes the code less transparent, leading to simulations that are difficult to modify and offer little possibility of reuse.

In an object-oriented paradigm, the coding objective emphasizes the *things* that make up the ecosystem (i.e., objects) and their interactions (Stroustrup, 1994) (e.g., in a vegetation model an object could be a tree or a life form hat interact with another tree or life form competing for water). The objects constitute the building blocks of a simulation model. Under this paradigm, a simulation consists of an assemblage of interacting objects. Each object in the simulation represents an object in the real world, having properties such as state and behavior, emphasizing realism in a simulation (e.g., the state of the object tree could be *height* and *amount of leaf carbon*, while its behavior could be modeled by actions such as radiation absorption, competition for light and photosynthesis). An object is a variable that is an instance of a class. A class is a user defined type which group together data (state) and functions (behavior). Functions are called methods in an object-oriented paradigm.

The object-oriented paradigm overcomes some of the disadvantages presented by the procedural approach. It provides mechanisms that enhance code reusability, such as the encapsulation of the details of implementation, the inheritance of properties from one class to another; and the ability to display polymorphic behavior by allowing related classes to respond differently to the same message. However, despite these mechanisms, it is not easy to reuse classes in different simulation models.

Component-based programming fosters a new programming paradigm that may be considered the next step after object-oriented and procedural programming (Troelsen, 2000). At first glance, a component-based approach is similar in many ways to object-oriented methods. In fact, the component-based programming paradigm builds on object-oriented concepts.

In a component-based programming paradigm simulations are conceived as an assemblage of interacting components. A component is built as an object that supports one or more interfaces. An interface is a set of operations that manipulates the class. For example, ModCom creates object-oriented simulations in which each component is a simulation object (simObj) implemented as a class (Hillyer et al., 2002). Components become the building blocks of a simulation model, where each component encapsulates software code representing objects or processes. The advantage here is that a component can be implemented internally as a single class, a set of interacting classes, or one or more procedures. Thus a simulation built under this paradigm can represent ecosystems, indistinctly, in terms of either processes or

things or a combination of the two. Another advantage is that components can employ existing code, facilitating the re-use of models.

In both component- and object-oriented paradigms, the object functionality is accessed through the interfaces that the object defines. However, components show a separation between the definition of the interface and its implementation, offering a deeper level of encapsulation of information. The implementation of the component is completely hidden and sometimes available only in binary form. Components created in this way are binary, self-contained, and language neutral. To use a component, there is no need to have knowledge of its implementation or its location; only the function of the component. As a result, component-based paradigms lead to robust design of simulation models, where it is easy to create new components, to re-use existing ones, and to swap components that access the same interfaces.

A component-based simulation framework is a generic software architecture that can be customized to create specific applications. These frameworks are typically composed of abstract and concrete classes and the interface between them. An abstract class is a class that specifies methods that have no implementation at all (these methods are called pure virtual functions) and that the modeler would need to implement. A concrete class is a class that is already implemented. Particular details of an application are implemented by adding components and by providing concrete implementations of abstract interfaces in the framework. A component is defined by its interfaces. The interfaces define the service the component provides and they also specify what services the component requires from the system that uses the component. When a system needs some service, it calls on a component to provide

that service without regard for where that component is executing or which programming language the component uses.

Incremental development of a simulation

When developing an ecological simulation, there are a number of different ways of organizing the activities of conceptualization, implementation and evaluation. Traditionally, these activities are organized sequentially (Figure 1), starting from the conceptualization, which begins with specifying the simulation requirements followed by an analysis and design phase to determine how the requirements will be achieved. Once the system is designed it is implemented and evaluated. Finally, the simulation is used and maintained. While this process results in well-structured simulations, they are inflexible and difficult to modify (Sommerville, 2001), and thus are inappropriate for the development of complex ecological simulations, such as those required by, for example, global vegetation models.

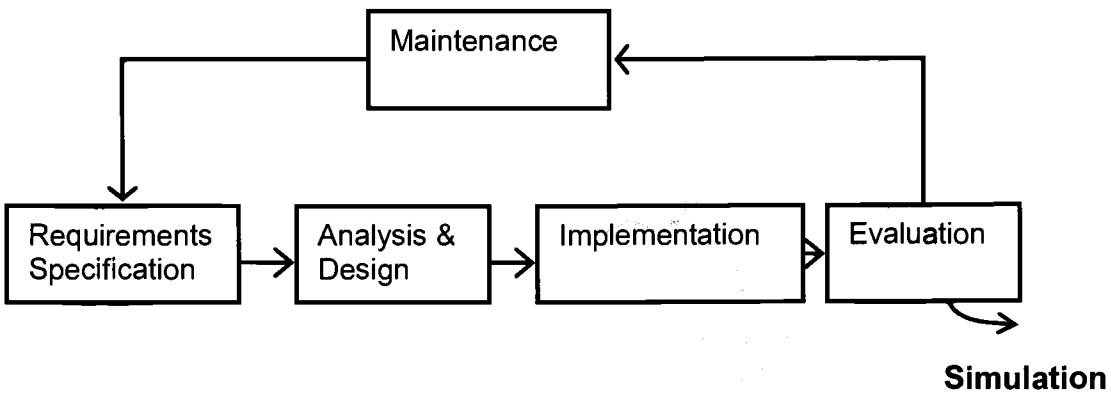


Figure 1. Sequential model of simulation development

In place of this sequential process, we propose an incremental process of simulation development that is more suitable for such complex ecological simulations. The process we describe is iterative, beginning with performance specification, which consists of establishing the scope of the simulation model, setting its constraints and its goals (Figure 2). A design phase follows, which involves identifying what subsystems are needed to achieve the goals of the simulation. At this point, the developer prioritizes the development of the most important subsystems and, based on that information, defines the increments contributing to the system's functionality (Sommerville, 2001). Once the necessary system increments are identified, the first increment is developed.

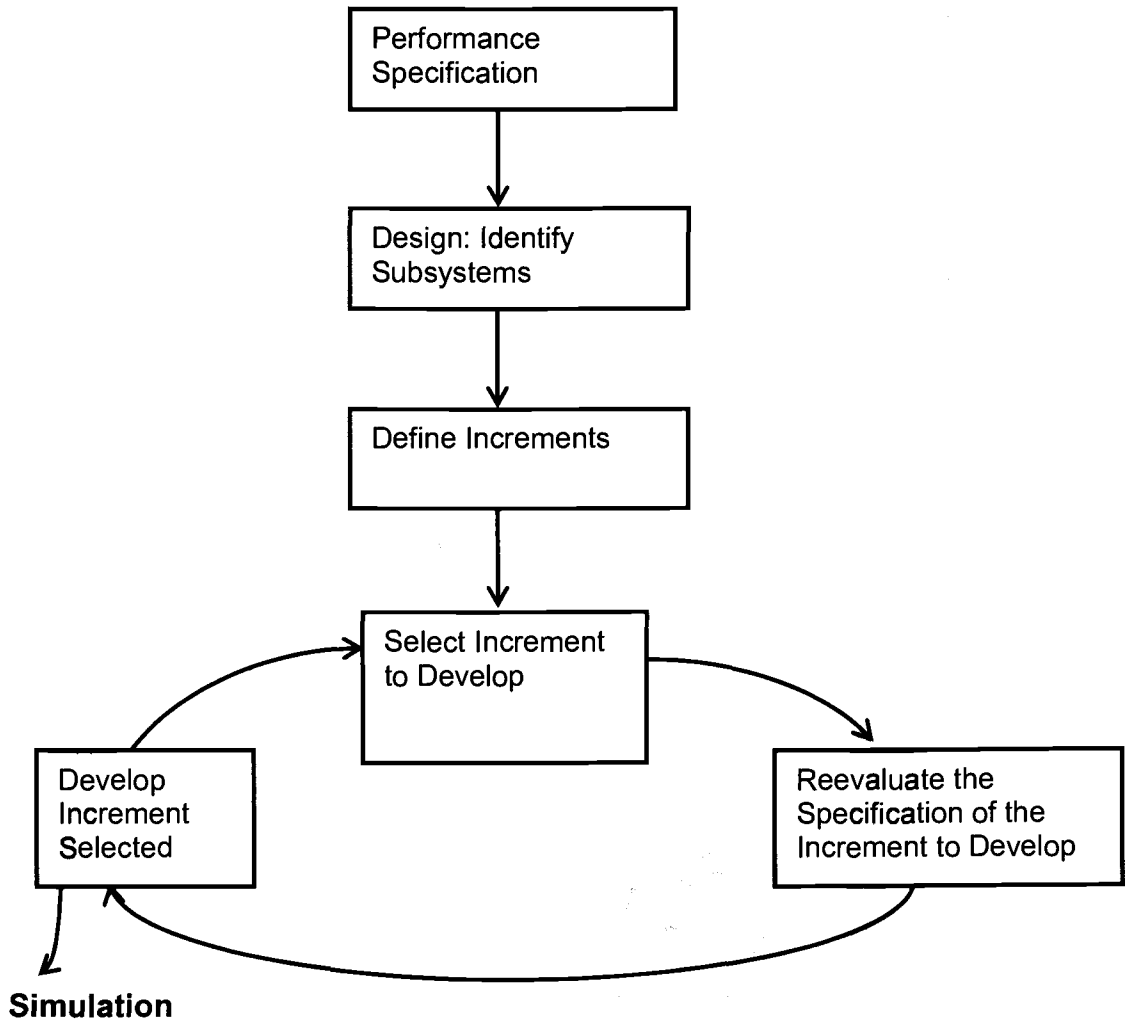


Figure 2. Incremental model of simulation development.

Developing an increment implies analyzing its requirements in detail and, if necessary, further decomposing it into subsystems. These subsystems are essentially service providers, whose services in a component-based paradigm, would be provided by the components. Once the services are defined in detail, the developer identifies the components that can provide those services and determines how these components will communicate with each other. Finally, these components are assembled in a simulation.

Increments are delivered one at a time. As each increment is added, the developer can test the simulation model and may decide to modify the requirements that still need to be completed. Finally, after the most recent simulation has been reevaluated, the next increment is implemented and the development and testing cycle is repeated. What guides the development of the simulation is the understanding of the modeled system (as enhanced by results from iterative testing of the model), rather than a preconceived notion about how the system should work.

In this paper we introduce and examine the efficacy of *component-based frameworks* in contrast with more popular approaches used in ecological modeling, such as process and object-oriented programming. The objectives of this paper are to evaluate the adaptability of component-based frameworks for:

1. effecting an incremental development of a new simulation and
2. modifying existing formulations so different modeling assumptions can be tested.

These questions are explored through a case study using the simulation framework ModCom (Hillyer et al., 2002).

Methods

Using the ModCom simulation framework, we explore the application and behavior of component-based frameworks through a case study. The utility of this novel technology is evaluated based on its support of incremental model development, and its ability to test different assumptions and to allow modification of existing model formulations.

ModCom

ModCom is a component-based simulation framework that can be extended to create a more specific application by assembling preexisting components and/or newly-created ones into an integrated simulation. In ModCom a component is a simulation object and simulations are conceived as a collection of simulation objects that are managed by a simulation environment. Simulation objects encapsulate simulation code that represents objects, processes, or existing models. These simulation objects are binary, self-contained and language neutral. The simulation environment provides general services that are common to all simulations (e.g., data analysis, data management, visualization). The modeler needs to define only the components of interest while all other details of the simulation are handled by the framework and are transparent to the modeler.

ModCom provides a series of interfaces that specify behavior supported by the simulation objects. *ISimEnv* and *ISimObj* are the two interfaces that are essential for ModCom (Figure 3). The *ISimEnv* interface provides high-level support for managing the simulation. This interface defines the functionality for (1) inter-object communication; (2) the registration and management of simulation objects participating in a simulation; and (3) the time-flow synchronization and control of the execution of updatable objects. The *ISimObj* interface defines basic behavior of the simulation object, such as how to expose variables (inputs and outputs) to communicate with other objects in the system. This interface is required for the objects that participate in the simulation. Other interfaces provide more specialized behaviors of simulation objects (e.g., *ISimData*, *ISimDataInfo*). Additionally, ModCom

defines other interfaces that provide statistical services, visualization and data management.

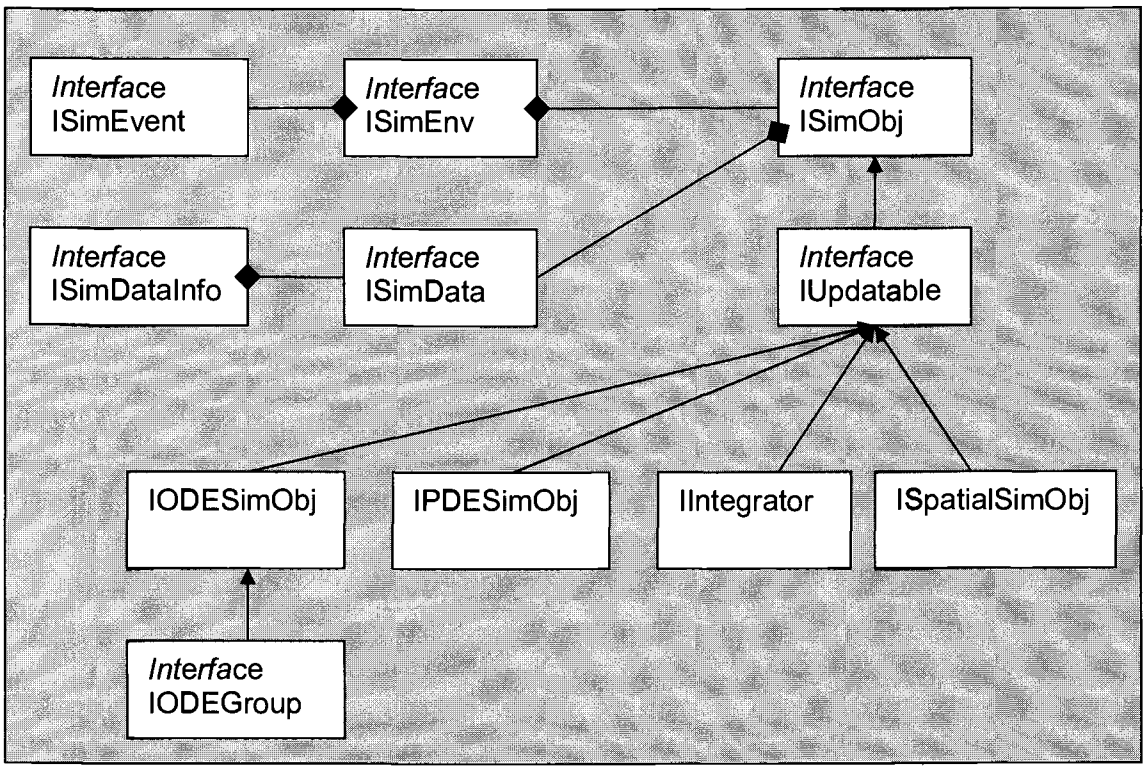


Figure 3. Core simulation services provided by ModCom.

Our case study consists of developing a simulation that calculates gross primary productivity (GPP) in a given ecosystem. In our example, we chose a forest ecosystem dominated by*species A* in the overstory and, in a more complex case, *species B* in the understory. Development starts with a very simple, one-species model, that estimates GPP from a linear relationship with light (aPAR) absorbed by the canopy (see Appendix). Later, this model evolves into a more complex formulation that takes into account the effects that vapor pressure deficit (VPD) and water limitations have on GPP. Adding a new level of complexity, the simulation is adapted to consider the concurrent presence of two life forms (e.g., trees and grasses). Finally, the simulation is modified to consider a new carbon allocation algorithm.

Results

The simulation was developed in four increments. The first increment implemented the simplest model, where GPP was estimated as a linear function of aPAR. The second increment implemented a more complex model, where GPP was constrained by VPD and water limitations. The third increment expanded the simulation to consider two life forms. The forth increment tested a different algorithm for the allocation of carbon between roots and leaves. Each increment was developed conceptually and then implemented physically in a simulation program. In all increments, the implementation code for the components was programmed in C++, and the components were assembled using Visual Basic. Data were input from a

spreadsheet program (Microsoft Excel), and output values were recorded and visualized using the same software.

The first increment was described in terms of the processes and the state variables needed for the estimation of aPAR to compute GPP (see Appendix). These processes include such activities as allocation of carbon into leaves and radiation uptake and the state variables represent the attributes of the objects subject to such actions (e.g., foliar mass). Processes modeled included the components: (1) *InputClimate*, which reads the climatic input data; (2) *RadiationUptake*, which computes the aPAR; (3) *GPPCalc*, which calculates gross primary productivity; (4) *CarbonAlloc*, which calculates the amount of carbon allocated into the leaves as foliar mass, and the amount of carbon lost as litterfall; and (5) *LifeFormState*, representing the lifeform's foliar mass and LAI (Figure 4). *LifeFormState* informs the amount of leaf carbon present in a lifeform after the allocation amount is added.

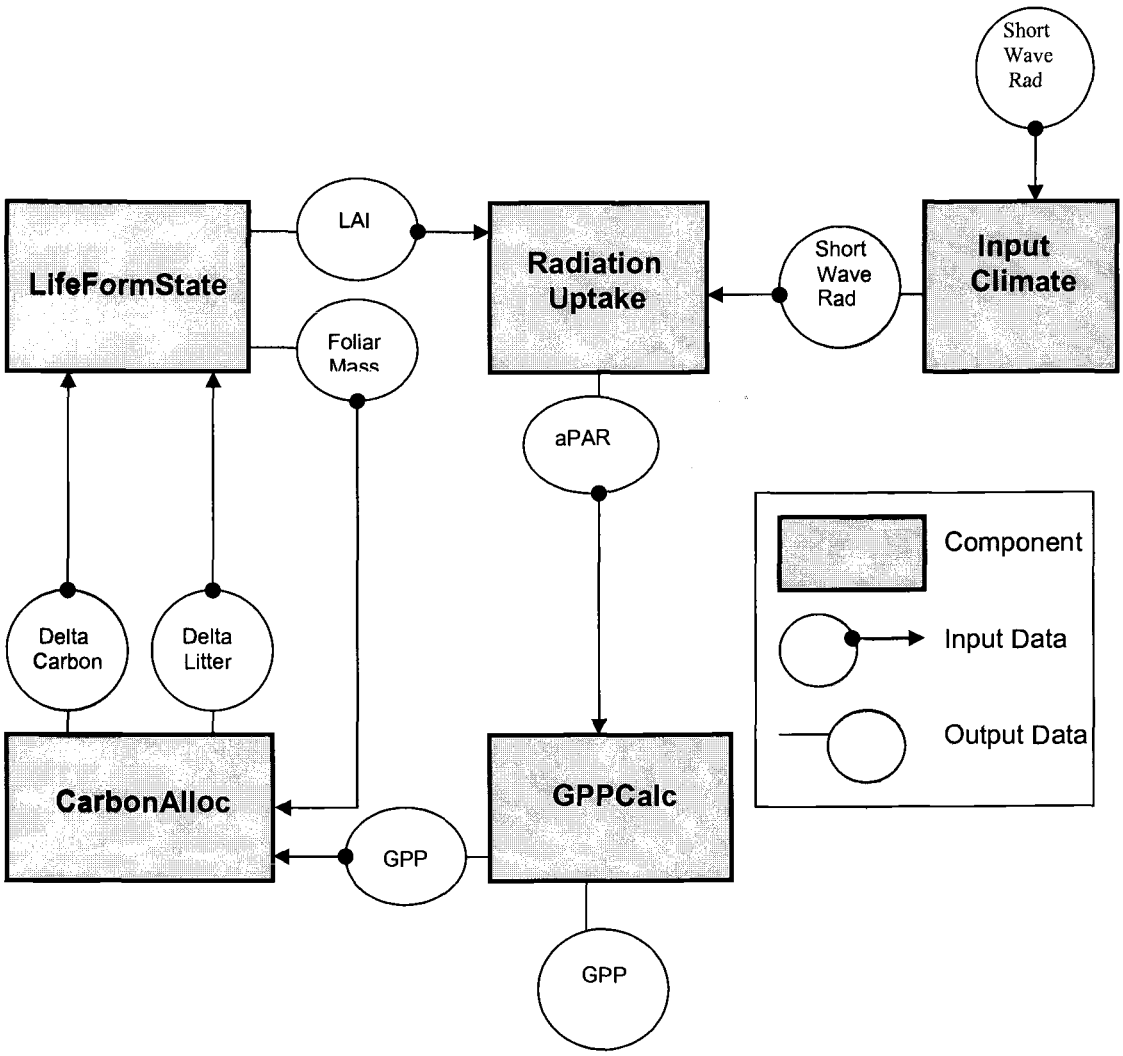


Figure 4. Components and their relationships in the carbon cycle

To implement the conceptual model in a simulation program, it was necessary to code the components as simulation objects, to determine how to pass information among simulation objects, and to assemble the components in a simulation program.

Implementing components. Each component identified during the conceptualization was implemented as a simulation object. The five simulation objects in this simulation implemented the *ISimObj* and *IUpdateable* interfaces. The *IUpdateable* interface

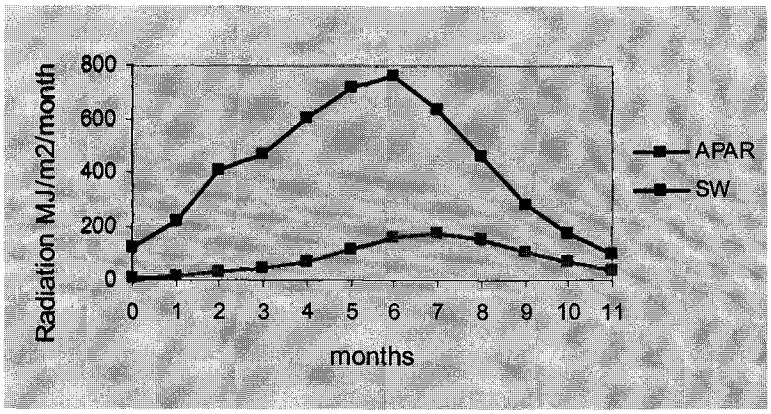
allows the objects to be changed when they are scheduled for update. For example, the simulation object *RadiationUptake* that calculates the aPAR defines two inputs (short-wave radiation and LAI) and one output (aPAR). In this component, the update method consists of calculating how much PAR is absorbed by the lifeform, given the short-wave radiation received as input and the life form's LAI. The five components were packaged as a single module in a dynamic link library (*dll*) named the Carbon Cycle.

Communicating and passing information among simulation objects. The simulation objects used the *ISimData* and *ISimDataInfo* interfaces for interobject communication. It was not necessary to implement these interfaces, since the default implementation provided by ModCom satisfied the exchange needs. A simulation object made information accessible to other simulation objects through these interfaces. A simulation object requested information from another simulation object by querying the object directly.

Assembling the simulation. Assembling the simulation consisted in: (1) creating an instance of the simulation environment; (2) creating instances of the participating simulation objects; (3) registering the simulation objects with the simulation environment (where registering means to make the simulation environment aware of the simulation objects that participate in the simulation); (4) establishing the communication between the simulation object's input/output pairs; and (5) calling the *Run* method so the simulation is executed. The simulation was managed through the *ISimEnv* interface provided by the simulation environment, which was responsible for initializing, scheduling and stopping the simulation.

The results of the first simulation show how GPP values respond to radiation, with higher GPP values observed during the summer months when the radiation is the highest (Figure 5b).

(a)



(b)

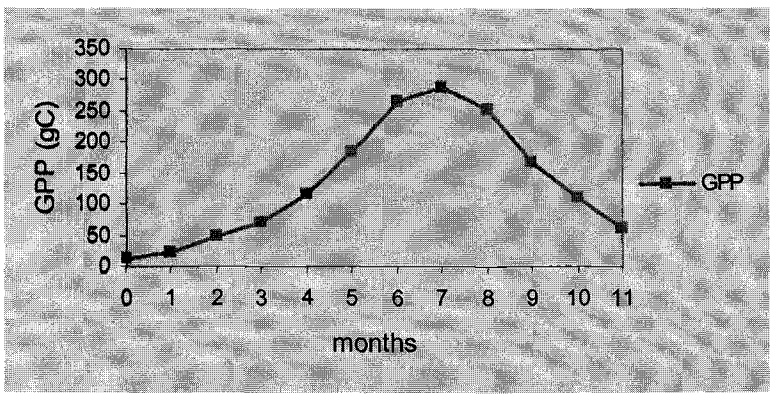


Figure 5. (a) input short-wave radiation and aPAR; (b) GPP. GPP is computed only as a function of aPAR.

Second increment

The development of the second increment of the model took into consideration the site water balance and the calculation of modifiers that characterized how water and VPD influenced the GPP estimate (see Appendix). The component *GPPCalc* from the first simulation was modified to account for the effects of water and VPD using a water and a VPD modifier. To incorporate these influences into the model, four additional components were required (Figure 6). Three of these components represented processes needed to compute the VPD and water multipliers: (1) *InputClimateWB*, which reads the climatic input data; (2) *WaterBalanceCalc*, which calculates the amount of water lost via evapotranspiration; and (3) *ModifiersCalc*, which calculates the water balance and VPD modifiers. A fourth component, *SoilWaterState*, represents the soil water content.

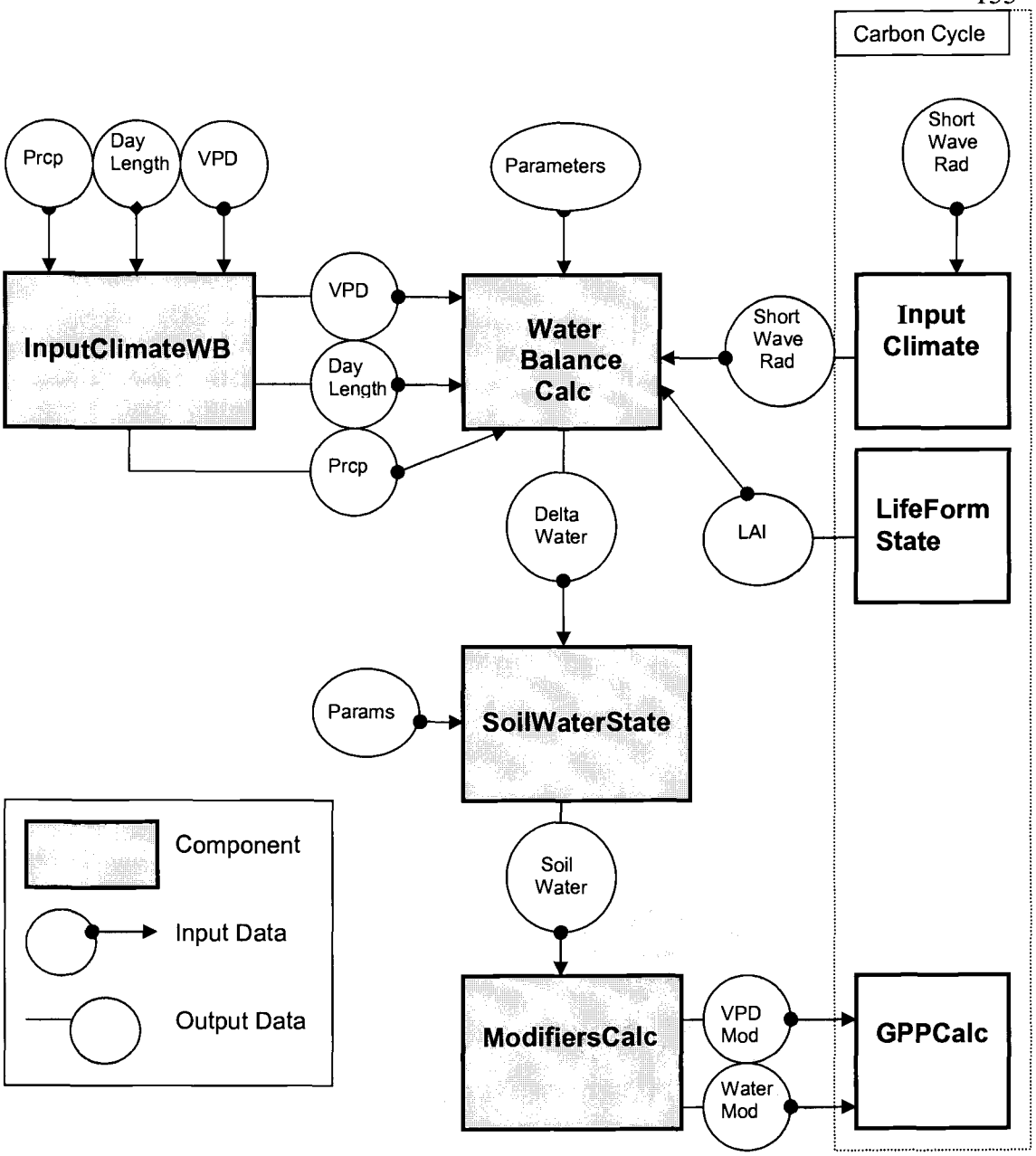
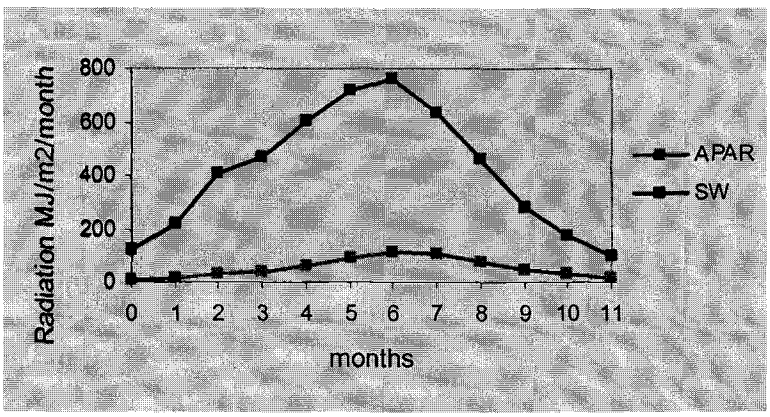


Figure 6. Structure of the second increment model. Components added to the simulation to compute the water and VPD multipliers are shaded blue. Non-shaded components belong to the carbon cycle. Not all components of the carbon cycle are shown in the diagram.

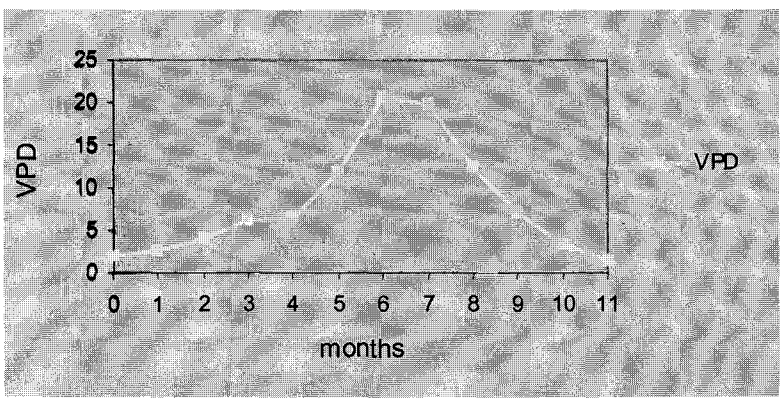
The implementation of the expanded conceptual model in a simulation program was carried out in the same way as in the first increment. The four components—

InputClimateWB, *WaterBalanceCalc*, *ModifiersCalc*, *SoilWaterState*—were packaged as a single module in a new *dll* named Water Cycle. In this case, the simulation objects that participated in the simulation belonged to two different dynamic link libraries, so assembling the simulation required integrating objects from both modules. In the second simulation, water and VPD constraints lowered GPP during the period simulated, particularly during summer months where VPD has the highest values and precipitation is limited (Figure 7a-d).

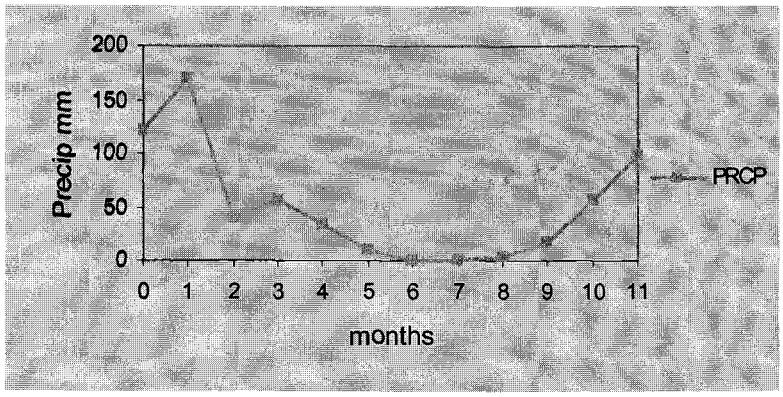
(a)



(b)



(c)



(d)

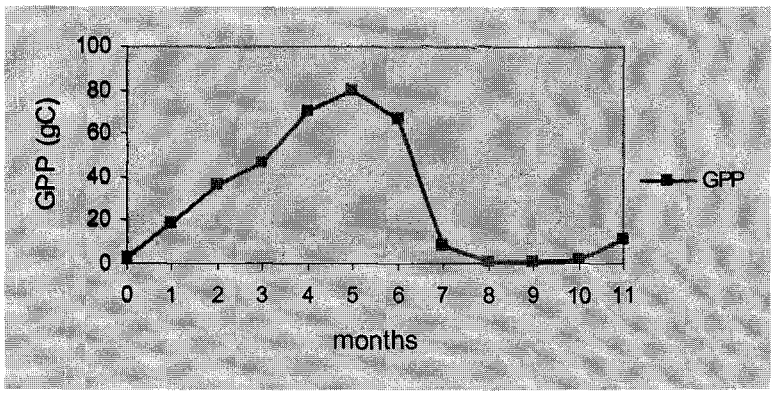


Figure 7. (a) input short-wave radiation and aPAR; (b) input VPD; (c) input precipitation; (d) GPP. GPP is computed as a function of aPAR constrained by VPD and water availability for one lifeform.

Third increment: Adding a second life form

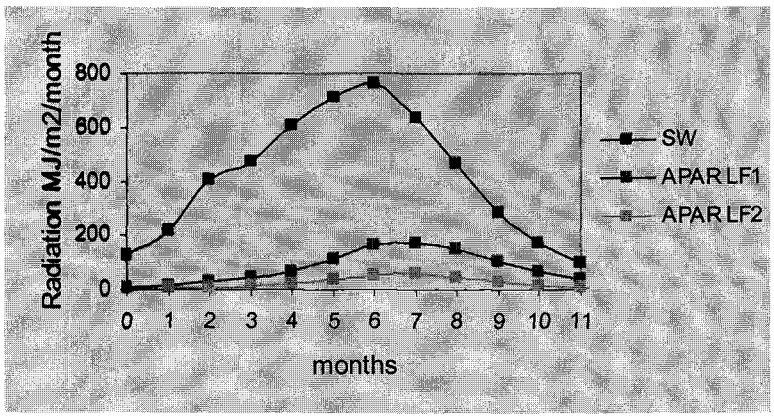
In the third increment of the model, a second life form was added. Allowing for two life forms required a series of modifications in the existing components. The radiation uptake (*RadiationUptake*) and water balance (*WaterBalanceCalc*) components were modified to account for radiation and water competition between the two life forms, overstory and understory (see Appendix). Other components (e.g., *GPPCalc*) did not require any modifications, so they were used just as developed in the first and second

simulations. The third increment was developed in two stages. First, GPP was modeled as a function of aPAR; in the second stage, the GPP calculation was constrained by water and VPD.

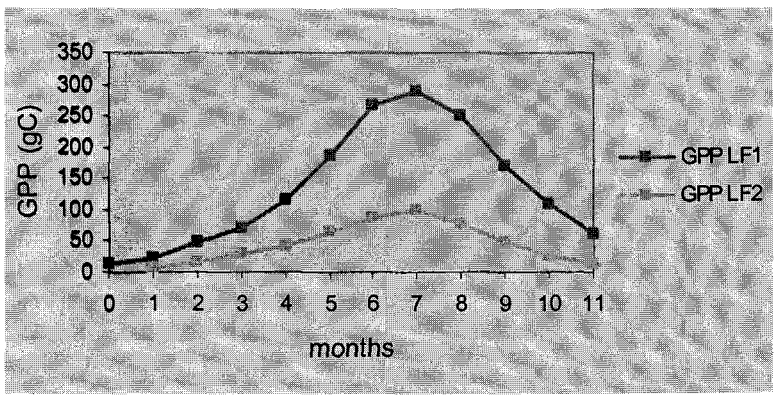
Modification of the *RadiationUptake* and *WaterBalanceCalc* components consisted of adding one more input that represented the second life form. These modifications also involved changing the Update method in the *IUpdateable* interface to simulate competition between the life forms. When assembling the simulation, the objects *StateVars*, *StateVarsWater*, *CarbonAlloc*, *GPPCalc* and *vpd_h2o_modif* were instantiated once for each life form. This simulation used the same climatic data as in the previous examples. The patterns of GPP obtained for life form (LF1) and life form (LF2) when GPP was computed as a linear function of aPAR did not change substantially (Figure 8a). As expected in this case, the overstory canopy resulted in the same values of GPP as in the first increment (Figure 5), while the understory canopy resulted in the smallest GPP of both lifeforms. This was because in this latter case, the light available for absorption was computed after the overstory canopy used all what it was needed.

In the case when GPP was constrained by water and VPD (Figure 8b) the dynamics were similar to those shown in the simulation resulted when the second increment was added (Figure 7), showing a decrease in GPP when water was scarce and VPD high. The understory canopy yielded smaller values of GPP.

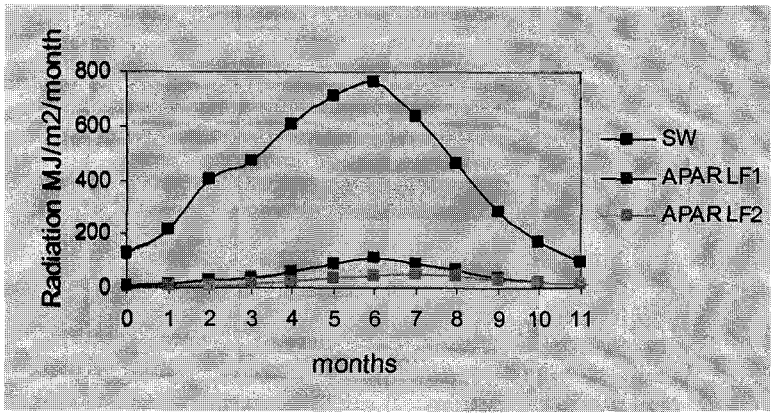
(a.1)-


Radiation MJ/m2/month
800
600
400
200
0
0 1 2 3 4 5 6 7 8 9 10 11
months
SW
APAR LF1
APAR LF2

(a.2)-


GPP (gC)
350
300
250
200
150
100
50
0
0 1 2 3 4 5 6 7 8 9 10 11
months
GPP LF1
GPP LF2

(b.1)-



(b.2)-

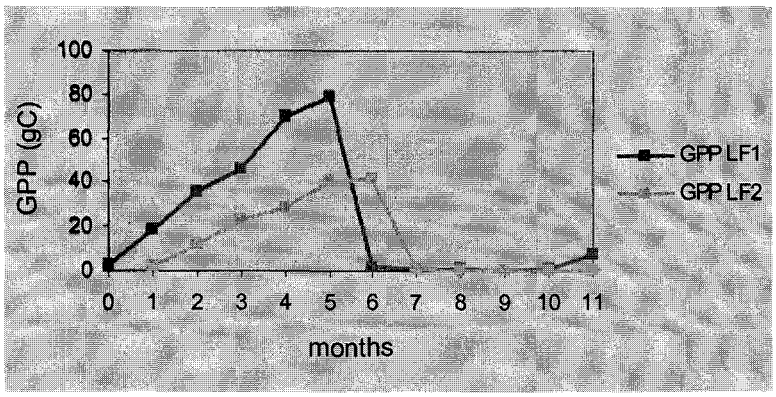


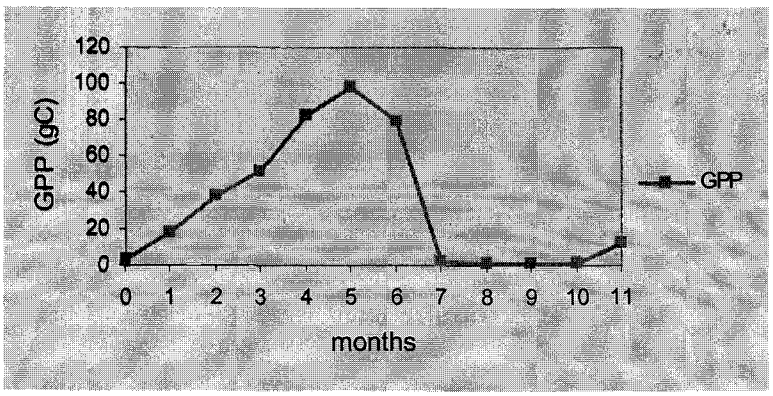
Figure 8: Radiation data and GPP for two life forms. (a.1) Input short-wave radiation and aPAR; (a.2) GPP. GPP is computed as a function only of aPAR. (b.1) Input short-wave radiation and aPAR; (b.2) GPP. GPP is computed as a function of aPAR constrained by VPD and water availability.

Fourth increment: Testing a new carbon allocation algorithm

In the fourth increment, a new carbon allocation algorithm was tested, in place of the old allocation scheme, in which carbon was allocated equally to leaves and roots under all conditions (Equation 6 in Appendix). The new carbon allocation model allowed the amount of carbon allocated to leaves and roots to be affected by growing conditions, favoring allocation to roots under harsh conditions (see Appendix). The

fourth simulation used the same climatic data as the previous three examples. When running the simulation with a single life form, the GPP values are higher to those obtained with the previous carbon allocation scheme (compare Figure 9a and Figure 7). However, when two life forms are considered, the understory life form (LF2) out-competes the overstory vegetation (LF1), resulting in higher GPP for the understory, as shown by higher values of GPP for LF2 (Figure 9b). Showing how different formulations of the carbon allocation process, which account for the uncertainty existing with the theory of this process, can yield to different results.

(a)-



(b)-

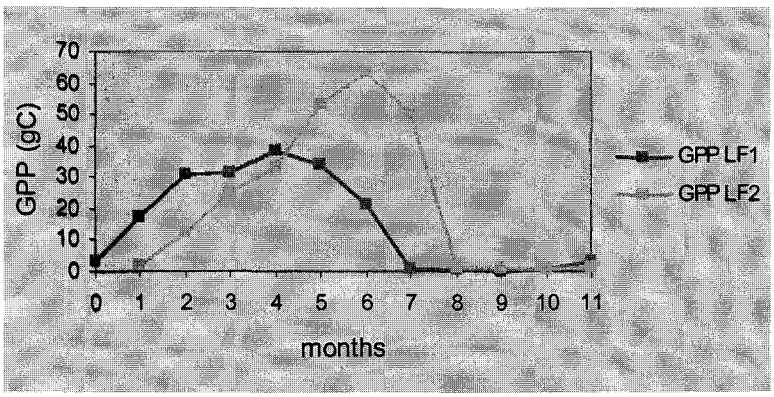


Figure 9: GPP under the new variable carbon allocation mechanism when: (a) GPP is computed as a function of aPAR constrained by VPD and water for one life form only; and (b) GPP is computed as a function of aPAR constrained by VPD and water for two life forms.

Discussion

Traditionally, the process of writing simulation software was a specialized technical task that produced simulation code for a specific purpose, code that was difficult to understand, to modify, and to reuse. These characteristics imposed severe limitations on what could be simulated. The introduction of component-based frameworks that support the development of ecological simulation opens new avenues for the

development of ecological applications of modeling. Component-based frameworks provide a programming environment in which it is possible to share code, to interchange ideas and techniques among simulation communities, to improve the structure of simulations, and to easily reuse previously written code. The flexibility provided by this new technology not only benefits model implementation, but it also indirectly influences the way in which a problem is conceptualized and evaluated.

One of the advantages that component based frameworks presents over procedural and object oriented paradigms is that they allow the easily incorporation of existing code into the simulation. Existing code can be incorporated as binary, self-contained, and language-neutral components facilitating the integration of existing models without the need to rewrite their code and to know their internal details. This characteristic represents a benefit to pursue an incremental development, since increments can be thought in terms of components that already exist, or in terms of components that can be developed independently (e.g., by different simulation communities).

One area of interest in the application of these concepts is in the development of simulation models from the coupling of already existing models. For example, dynamic global vegetation models that are built by coupling biogeochemical with biogeographical models, or larger environmental simulations built from coupling land, atmosphere and ocean models. As a result, simulations can be built more rapidly, taking advantage of already existing resources and bringing together the expertise of interdisciplinary teams of scientists.

Nevertheless, the modeler needs to be aware that even though the incorporation of a component only requires knowing what the component does regardless to its internal details, evaluating the simulation requires taking into account how the component works and which are the assumptions built into it. Also the modeler has to be cautious about the incorporation of components that are only available in binary form, because there are no possibilities of modifying that particular piece of code. This may create problems if the addition of a new increment of the model requires the modification of such binary components.

For example, in our case study, the incorporation of the components that implemented the water balance, developed in the second increment, required modifying the calculation of GPP to account for the water and VPD multipliers. This involved modifying the component that calculated GPP (*GPPcalc*) so it could accept two more inputs (the water and VPD multipliers), as well as modifying the piece of code in which the actual value for GPP was derived. These modifications were feasible only because the code that calculated GPP was available. If the code was in binary form these modifications would not have been possible.

In our model simulation example, we found that component-based simulation frameworks were somewhat limited in their ability to modify existing simulations. Although they offered certain benefits, their disadvantages were similar to those presented by other approaches. Among their benefits, component-based frameworks generate robust code, where the use of a component is independent from its implementation. So when the internal implementation of a component changes, as occurred when we changed carbon allocation algorithms, there is no need to rewrite

or re-compile the code that used that component. In this regard, incremental simulation development provides additional benefits, where each subsystem is implemented as a well-related group of components, facilitating the modifications of the model subsystems that were confined to a particular increment.

The incorporation of new state variables, as carried out during the development of the third increment, required modification of both the carbon and water balance subsystems. These modifications were facilitated by grouping state variables representing the life form in a single independent component. Thus, the addition of a second life form (grasses) required only that the program create an instance of the state variable component twice; which implied using a different set of parameter values for the second life form, but the same component. The same rationale was followed for the carbon allocation and GPP calculation components. The processes that needed to be modified were the calculation of radiation uptake to model radiation competition and the calculation of the water balance to model water competition between overstory and understory vegetation. In addition, the code that assembled the simulation underwent major modifications. As would have happened in an object-oriented approach, to make the necessary changes to add state variables required an extensive knowledge about the structure and function of the components, and the availability of the code that needed to be modified.

The use of component-based frameworks brings new possibilities for the treatment of uncertainty in ecological simulations models. One benefit they offer is that they facilitate the investigation of the effects that different assumptions or hypotheses have on model behavior. So the modeler can evaluate the consequences of modeling a

problem from different perspectives. This benefit is enhanced when the simulation is developed incrementally, since the development is guided by an iterative testing of the system and the modeler does not need to commit to a single set of specification and design decisions before developing the simulation. As a result, the modeler can explore different alternatives of representation.

Unlike other approaches, testing different assumptions and hypotheses are done by coding each of them as a component that is alternatively assembled in the simulation. Doing so only requires the alternative components to expose the same variables without being necessary to know the internal details of the component, the programming language in which it is implemented, or its location. In this regard, a component-based approach presents several advantages with respect to other programming paradigms. In simulations that follow procedural programming, testing different assumptions or hypotheses is cumbersome. Since in a procedural program, the data and functions that logically belong together are not physically grouped together in the code, incorporating a different assumption or hypothesis in the code requires the developer to have a deep understanding of all the functions and global variables in the simulation. Programming the changed assumptions may necessitate modifications that may expand throughout the whole simulation code. In this context, attempting to link already existing code is a cumbersome activity. Object-oriented programs offer better possibilities. Since in these programs, there is a tight coupling between data and functionality, the developer can easily change a function without the need to understand the whole simulation. In addition, through the mechanism of polymorphism, the developer can easily implement different behaviors for a class, each one representing a different assumption or hypothesis. However, object-

oriented programming is still limited in its ability to link alternative existing code. To establish such a link, it is not only necessary to have knowledge of the structure of the simulation, to be certain that the new alternative is in agreement, but the code must also be written in the same language as the rest of the simulation in order to be incorporated

The use of component-based simulation frameworks enhances the correspondence between the way the problem is conceptualized and the way it is implemented, so that the conceptualization is clearly reflected in the resulting code. During conceptualization in the design phase of our example, the structure of the system was considered hierarchical, and each level was further decomposed into subsystems. This process was repeated until the subsystems could be easily understood, described and solved. Each subsystem was conceived as a provider of services to other subsystems that were needed to achieve the goals of the simulation. Each of these subsystems was implemented as a component, which was grouped with other components belonging to the same subsystem. This structure mirrored the hierarchical organization expressed during the model conceptualization. Since development was carried on in an incremental fashion, the conceptualization of the system was interwoven with the implementation.

An implementation approach that mirrors problem conceptualization leads to simulations that are easy to understand. Because assumptions and simplifications involved in conceptualization of the ecosystem are reflected in the code implementation, the investigation of their effects on the predictions is simplified. This improves the communication between programmers and scientists, facilitating their

co-participation in the process of development. Additional benefits accrue when the implementation is comprised of components that can be built independently. This facilitates the participation of specialists in the development, who can use their expertise to build or modify a particular component without the need to understand how the rest of the simulation works.

Conclusions

In our view, frameworks constitute the next step in programming paradigms, which by providing a common infrastructure for the development of ecological simulations and promoting the use of existing resources, make simulations models easy and rapid to develop. The use of component-based frameworks raises important questions about the state and direction of ecological modeling. Unlike previous paradigms, they better manage the investigation of uncertainty since assumptions, hypothesis and simplifications built in a simulation can better be evaluated. In this way, they promote the investigation of the many options and alternatives required to understand how an ecosystem works and how it can be simulated, shifting the modeling activity from being focused on pursuing a singular representation of the ecosystem to allow for a plurality of possible explanations. Overall, we consider that this new technology provides a programming paradigm that better mirror the development of ecosystem theory and that can accommodate to the progressive changes in knowledge and understanding of ecosystems.

References

Allen, T.F.H. and T.W. Hoekstra. 1992. Toward a Unified Ecology. Columbia University Press, New York, 384 pp.

Baumann, M. 2000. On nature, models and simplicity. Conservation ecology 4(2): r4.[online] URL: http://www.consecol.org/vol4/iss2/resp4.

Flynt, J. and J. Mauldin. 2001. An example specification for implementing a temperature Regulator software component. *In* Component-based software engineering. Heineman, G.T. and W.T. Council. Editors. Addison Wesley, Usa, pp 49-69.

Farquhar, G.D., S. von Caemmerer and J.A. Berry. 1989. A biochemical model of photosynthetics CO2 assimilation in leaves of C3 species. Planta 149:79-90.

Hillyer, C., J. Bolte, F. van Evert and A. Lamaker. 2002. The ModCom Modular Simulation System. Journal of European Agronomy. In press.

Houston, K. and D. Norris. 2001. Software components and the UML. *In* Component-based Software Engineering. Heineman, G.T. and W.T. Council, Editors. Addison Wesley, Usa, pp 243-262.

Jarvis, P.G. and J.W. Leverenz. 1983. Productivity of temperate, deciduous, and evergreen forest. In encyclopedia of plant physiology. O.L. Lange, P.S.Nobel, C.B.Osmond, and H.Zeigler, eds. Springer-Verlag, New York, pp.233-280.

Klir, G.J. 1991. Facets of Systems Science. Plenum Press, New York, 664 pp.

Klir, G.J. and M.J. Wierman. 1999. Uncertainty-based information. Elements of generalized information theory. Physica-Verlag, 168 pp.

Landsberg, J.J. 1986. Physiological Ecology of Forest Production. Academic Press. London, 198 pp.

Landsberg, J.J. and R.H. Waring. 1997. A generalized model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. Forest Ecology and Management, 95:209-228.

Running, S.W., and J.C. Coughlan. 1988. A general model of forest ecosystem processes for regional applications. I Hydrological balance, canopy gas exchange and primary production processes. Ecological Modeling. 42:125-154.

Sommerville, I. 2001. Software Engineering. 6th Edition. Addison-Wesley, Usa, 693 pp.

Stroustrup, B. 1994. The C++ Programming Language. Addison-Wesley, Usa, 691 pp.

Thornton, P.E. 1998. Description of a numerical simulation model for predicting the dynamics of energy, water, carbon and nitrogen in a terrestrial ecosystem. Ph.D. dissertation. University of Montana, Missoula, MT 280 pp. [Available from Mansfield Library, University of Montana Missoula, MT 59812].

Troelsen, A.W. 2000. Developer's workshop to COM and ATL 3.0. Wordware Publishing, Inc., Texas, 697 pp.

Zeigler, B.P. 1984. Theory of modelling and simulation. Krieger Publishing Company,Malabar, Florida, pp 435.

Zimmermann, H.-J. 2000. An application oriented view of modeling uncertainty. European Journal of Operational Research, 122:190-198.

Appendix

First Increment

In the first increment, GPP is calculated using a linear relation with the light absorbed by the canopy, as described in Equation 1.

$$GPP = \alpha * aPAR \quad (1)$$

where,

α is quantum efficiency, and

aPAR is the absorbed photosynthetic active radiation.

The model requires monthly values of incoming short wave radiation (I_{sr}) as climatic information. These data are used to estimate the photosynthetic active radiation (PAR), as indicated in Equation 2. The aPAR is computed by applying the Beer-Lambert's law to calculate the fraction of PAR intercepted by the layers of leaves in the canopy, as described in Equation 3.

$$PAR = 0.5 * Isr \quad (2)$$

$$aPAR = PAR * (1-\exp(-k * LAI)) \quad (3)$$

where,

LAI represents leaf area index,

-k is the extinction coefficient, and

I_{sr}, denotes short wave radiation.

LAI is calculated by multiplying the foliar mass by the specific leaf area (SLA, m^2/kg; Equation 4). Every month, the foliar mass is incremented and new carbon is allocated into foliage. The carbon available for growth is specified by the net primary productivity (NPP), which is the result of GPP minus respiration. In this model, the ratio NPP/GPP is conservative (Equation 5) and the carbon allocated to leaves is a fixed portion of the NPP (Equation 6). The model considers litterfall, which represents leaf carbon lost every month. Litterfall is computed as indicated in Equations 7 and 8, following the same rationale as in Landsberg and Waring (1997).

LAI = Foliar mass * SLA (4)

NPP = GPP * 0.46 (5)

Foliar mass = NPP * 0.37 (6)

Litterfall = FoliarMass * Littfall_rate (7)

Litterfall_rate = MaxLitterRate * InitialLitterRate /

{InitialLitterRate + (MaxLitterRate - InitialLitterRate) *

exp[-12 * log(1 + MaxLitterRate / InitialLitterRate) * month]} (8)

where,

MaxLitterRate is defined as the maximum litterfall rate, and

InitialLitterRate is the initial litterfall rate.

Second Increment

The second increment develops a more complex formulation of the model, which consists of constraining the GPP calculation with the effects of vapor pressure deficit (VPD) and water limitations. The model is adjusted to include processes that compute water balance (i.e., precipitation interception, transpiration and soil water content), and a new state variable that represents the soil water content. These changes imply new requirements for input data, such as monthly values of precipitation and VPD. GPP is then calculated as shown in Equation 9.

$$GPP = \alpha * aPAR * minimum(f\_H2O, f\_VPD) \quad (9)$$

where,

f_H2O is the water multiplier, and

f_VPD is the VPD multiplier.

Both f_H2O and f_VPD have a value between 0 and 1. The multipliers are calculated in the same way as in Landsberg and Waring (1997; Equations 10 and 17). The water balance used to estimate the Moisture Ratio in Equation 10 is computed in Equations 11 through 16.

f_H2O = 1 / {1 + [(1 - Moisture Ratio) / SoilWaterConst]SoilWaterPower} (10)

where,

SoilWaterConst and SoilWaterPower are parameters that vary according to soil type.

Moisture Ratio = Soil Water Content / Maximum Soil Water (11)

Soil Water Content = Soil Water Content + Water Balance (12)

Water Balance = Precipitation - Evaporation-Transpiration (13)

Evaporation-transpiration = Transpiration + Interception (14)

Interception = Coeff_of_Intercept * Precipitation (15)

where,

Transpiration is computed following the Penman-Monteith equation, and

Coeff_of_Intercept is a coefficient of interception.

In this model, soil water content must be greater than a minimum soil water content, which varies with soil type.

f_VPD = exp(-CoeffCond * VPD); (16)

where,

CoeffCond is the stomatal response to VPD.

Third Increment

The third increment implements the most complex formulation, in which two life forms coexist. In this model, one life form must represent forest overstory plants, and the other understory plants. Competition for radiation, both short wave and PAR, is modeled by computing first the fraction of radiation intercepted by the overstory canopy, following the Beer-Lambert's law as described in Equation 3. Then, the radiation that reaches the understory is estimated by subtracting the radiation intercepted by the overstory from the total radiation received. Finally, the radiation that reaches the understory is used to compute the radiation absorbed by the understory canopy following the Beer-Lambert's law. Equations 17 through 20 describe this process for PAR. The same rationale is followed to compute short wave radiation interception.

$$TotalPAR = 0.5 * I_{sr} \quad (17)$$

$$aPAR\_overstory = TotalPAR * (1-exp(-k * LAI\_overstory)) \quad (18)$$

$$PAR\_understory = TotalPAR - aPAR\_overstory \quad (19)$$

$$aPAR\_understory = PAR\_understory * (1-exp(-k * LAI\_understory)) \quad (20)$$

Interception of water in the presence of two life forms is modeled by computing the amount of water intercepted by the overstory (Equation 15). This amount is subtracted from total precipitation to estimate the amount of water reaching the understory. The amount of water intercepted and used by the understory layer is computed in the same manner as that for the overstory (Equation 15).

Water balance is modeled by computing the amount of water that the life forms can potentially transpire. This amount is computed independently for each lifeform, and is used to estimate total transpiration at the site, which then is subtracted from the soil water content. If the soil water content is less than a minimum prescribed according to soil type, soil water content is set by default to its minimum amount.

Fourth Increment

The fourth increment implements a new carbon allocation algorithm. In the previous increments, the carbon allocation mechanism used assumed that carbon was allocated equally to roots and leaves under all conditions. The new carbon allocation algorithm used in this increment allows carbon to be allocated to roots and leaves *in* accordance to growing conditions. This allocation scheme is equivalent to the one presented by Landsberg and Waring (1997), and is described in Equations 21 through 23.

Utilizable aPAR = aPAR * minimum(f_H2O and f_VPD) (21)

Allocation to Roots = MaxRootAlloc * MinRootAlloc/

[MinRootAlloc + (MaxRootAlloc - MinRootAlloc)

* (Utilizable aPAR / aPAR)] (22)

Allocation to Leaves = (1 - Allocation to Roots) * LeafPartition (23)

where,

MaxRootAlloc is the maximum fraction of NPP allocated to roots,

MinRootAlloc is the minimum fraction of NPP allocated to roots, and

LeafPartition is the portion of above ground carbon allocated to leaves.

Chapter 5

Conclusions

I have built this dissertation around the concept of uncertainty in ecological models with two central objectives: (1) examining the problem and concept of uncertainty form a post-normal view; (2) investigating innovative methods for evaluating and identifying uncertainty within the post-normal framework, the idea being is that complexity theory and post-normal science have shifted the problem of uncertainty from the goal of its elimination to it identification relative to the problem and objective in hand. Here I suggest a post normal approach of modeling is more adequate than a normal one to model complex ecological problems. A post normal approach emphasizes the preponderant role of uncertainty in these models; recognizing that uncertainty is intrinsic to modeling and that dealing with it is unavoidable. It departs from its former negative identity where uncertainty is undesirable to the present case, where uncertainty is considered as a source of information to point out the limitations and subjectivities embedded in the model; acknowledging the capabilities of the model and the scope of its inference. In this context, models are not so much conceived as predictive devices able to foresee the future, but as tools that allow understanding a problem, embracing uncertainty and ignorance and welcoming a plurality of perspectives. Thus the study of uncertainty has told us as much about the characterization of uncertainty as it does about the nature and goal of models themselves.

Under this philosophical framework I examined uncertainty in rule-based classification and process-based models. In chapter 2, I introduced a methodology to evaluate the significance of model uncertainty in rule-based classification models. This methodology analyzed model behavior by studying how the model reacted to perturbations in threshold values. It also determined the most likely alternative output

classes that could be generated by the model when uncertainty in the threshold values was considered. It further provided the analyst a measure of model confidence that allowed interpreting the results. This measure of confidence conveyed information about the changes in the classification occurred when threshold values were varied, in conjunction with, information about the location of these changes in the classification (e.g., a change in a core area was considered more severe than one in a boundary area).

Overall the results from this methodology informed about the deficiencies in model predictions (e.g., vegetation classes miss-represented) and how significant these deficiencies were, considering its ecological and spatial context in which they occurred. This methodology highlighted the areas in which the classification was ambiguous and for those areas it provided alternative results. Besides, the confidence measure was an additional source of information to help evaluate to what degree the results of the model could be trusted. From a model evaluation perspective, the information provided by this methodology could be used to identify flaws in model logic and to determine what type of improvements may be needed. From a model interpretation perspective, this methodology provides the scientist and decision maker with valuable information about the quality and limits of model predictions. This methodology also provides a conceptual framework for comprehensively evaluating and interpreting modeling results where model capabilities and model limitations were both acknowledged. While this methodology only considers uncertainty in threshold values, future research should be directed to investigate other sources of uncertainty that may also affect model results. This study

brought a concrete approach to representing both what the modeler knows and what it is unknown or uncertain.

In chapter 3, I introduced a methodology that uses abstraction at the process level to perform sensitivity analysis. This method focuses on understanding how variations in processes affect the internal working of the model and the range of its predictions at the system level. Overall, this analysis was designed to address the critical question of a model's sensitivity to assumptions in processes and model structure. This analysis was conceived under the premise that a good modeling practice does not only entail building a model able to reproduce observable data, but it also requires building a model able to capture the structure of the real ecosystem and understanding how sensitive a model is to different assumptions.

The information provided by this methodology allows the modeler to corroborate whether the way in which the model operates is congruent with the understanding s/he has about the system being modeled. Additionally, the use of abstraction at the process level allows handling high levels of complexity by reducing the search space to a few processes instead of dealing with numerous parameters. It also facilitates the interpretation of model behavior by providing the information at the process level, claiming that it is easier to think in terms of a limited number of processes rather than in terms of many individual parameters, state variables or input data. Since excess of detail may hinder a modeler's abilities to interpret how a model behaves, a process level abstraction provides a more convenient and manageable way of dealing with model information. Besides, it is coherent with the fact that it is easier to understand model behavior when there is a correspondence between what it is represented in the

model and the real phenomena, than when there is no such a mapping between model and reality. This is the case of modeled processes that are conceptually related with real ecosystem processes as opposed to individual parameters or state variables that do not necessary have a real counterpart. For example, in the case study presented, when trying to understand what a model does, it is easier to think in terms of Maintenance Respiration as the action of computing the plant's metabolic rate, than to think in terms of the individual parameters q10 and mpern.

Finally, in chapter 4 I claim that current programming paradigms fail in offering the adequate programming environment to deal with the different sources of uncertainty in ecological models. To this end, in the third paper I investigated the programming possibilities component base simulation frameworks, such as ModCom, offers. These frameworks, while fostering a component based programming paradigm, promote software development by dividing a large project into discrete and manageable subprojects. I considered that they constitute a better option for the development of ecological simulation models. Frameworks provide a flexible environment for the testing of different hypothesis, for adapting to changes in requirements or specifications, for allowing for the interdisciplinary participation of experts, and for reusing code that already exist. These are characteristics that facilitate the investigation of uncertainty.

In my opinion future research should be directed towards the formulation of modeling tools that better mimic the process of modeling in which modelers and scientist engage; tools that can easily accommodate the progressive changes in knowledge and understanding of ecosystems while recognizing the subjectivities and limitations associated with the modeling process. The methodologies presented in Chapter 2

and 3 constitute an advance to this end, providing strategies to address some of the conceptual problems rule and process-based models present. Chapter 4 also provides an advance in this direction by investigating new implementation strategies that can better address these problems.

BIBLIOGRAPHY

Allen, T.F.H. and T.W. Hoekstra. 1992. Toward a Unified Ecology. Columbia University Press, New York, 384 pp.

Allen, T.F.H, J.A. Tainter and T.W. Hoekstra. 1999. Supply-side sustainability: Systems Research and Behavioral Science, 16:404–427.

Allen, T.F.H., J.A. Tainter, J.C. Pirex and T.W. Hoekstra. 2000. Dragnet ecology- "Just the facts, Ma'am: The privilege of science in a postmodern world. BioScience. 51(6):475-485.

Barbour, M.,G. L.H. Burk and W.D. Pitts. 1987. Terrestrial plant ecology. Second edition. The Benjamin/Cummings Publishing Company Inc., 634 pp.

Baumann, M. 2000. On nature, models and simplicity. Conservation ecology 4(2): r4.[online] URL: http://www.consecol.org/vol4/iss2/resp4.

Beck, B. 2002. Model evaluation and performance. Encyclopedia of environmetrics. In El-Shaarawi, A.H. and W.W. Piergorsch. John Wiley & Sons, Ltd, Chichester, 3:1275-1279.

Bradshaw, G.A. and J. G. Borchers. 2000. Uncertainty as information: narrowing the science-policy gap. Conservation Ecology, 4(1):7. [online] URL://www.consecol.org/vol4/iss1/art7

Campolongo F., A. Saltelli, T. Sorensen and S. Tarantola. 2000. Hitchhiker's Guide to Sensitivity Analysis. *In* Sensitivity Analysis, Edited by A. Saltelli, K. Chan and E.M.Scott. Willey Series in probability and statistics, pp 15-47.

Cartwright, H.M. 1995. The genetic algorithm in Science. Pestic. Sci., 45:171-178.

Chargaff, E. 1997. In dispraise of reductionism. BioScience. 47(11): 795-797.

Cilliers, P. 1998. Complexity and postmodernism. Understanding complex systems. Routledge, 154 pp.

Elston D. A. 1992. Sensitivity analysis in the presence of correlated parameter estimates. Ecological Modelling. 64:11-22.
Farquhar, G.D., S. von Caemmerer and J.A. Berry. 1989. A biochemical model of photosynthetics CO2 assimilation in leaves of C3 species. Planta 149:79-90.

Fishwick, P.A. 1988. The Role of Process Abstraction in Simulation. IEEE Transactions on systems, man and cybernetics, 18(1):18-38.

Flynt, J. and J. Mauldin. 2001. An example specification for implementing a temperature Regulator software component. *In* Component-based software engineering. Heineman, G.T. and W.T. Council. Editors. Addison Wesley, Usa, pp 49-69.

Forrest, S. 1993. Genetic algorithms. Science, 261:872-878.

Funtowicz, S. O. and J. R. Ravetz. 1990. Uncertainty and quality in science for policy. Kluwer Academic, Dordrecht, The Netherlands, 232 pp.

Funtowicz, S. O. and J. R. Ravetz. 1993. Science for the post-normal age. Futures. 25(7): 739-755.

Gardner, R. H., V. H. Dale and R. V. O'Neill. 1990. Error propagation and uncertainty in process modeling. In Process Modeling of Forest Growth Responses to Environmental Stress. Eds. R.K. Dixon, R.S. Meldahl, G.A. Ruark and W.G.Warren. Timber Press, Portland, OR, pp 208-219.

Green, E.J., MacFarlane, D.W. and H.T.Valentine. 2000. Bayesian synthesis for quantifying predictions from process models. Tree Physiology, 20:415-419.

Haefner, J. W. 1996. Modeling Biological Systems, principles and applications. Chapman and Hall, 473 pp.

Hann, D.W.,A.S. Hester and C.L. Olsen. 1997. ORGANON user's manual: Edition 6.0. Department of Forest Resources, Oregon State University, Corvallis, OR, 133 pp.

Hillyer, C., J. Bolte, F. van Evert and A. Lamaker. 2002. The ModCom Modular Simulation System. Journal of European Agronomy. In press.

Houston, K. and D. Norris. 2001. Software components and the UML. *In* Component-based Software Engineering. Heineman, G.T. and W.T. Council, Editors. Addison Wesley, Usa, pp 243-262.

Holland, J.H. 1992. Genetic algorithms. Scientific American, 267 (1): 66-72.

Jarvis, P.G. and J.W. Leverenz. 1983. Productivity of temperate, deciduous, and evergreen forest. In encyclopedia of plant physiology. O.L. Lange, P.S.Nobel, C.B.Osmond, and H.Zeigler, eds. Springer-Verlag, New York, pp.233-280.

Kann, A. and J.P. Weyant. 2000. Approaches for performing uncertainty analysis in large scale energy/economic policy models. Environmental Modeling and Assessment. 5(2000) 29-46.

Katz, R. W. 1999. Techniques for estimating uncertainty in Climate Change scenarios and Impact studies. In: Representing Uncertainty in Climate Change Scenarios and Impact Studies. Proceeding of the EU Concerted Action Initiative ECLAT-2 Workshop 1, Helsinki, Finland, April 16th - 18th, 1999, Climatic Research Unit, Norwich, UK., pp 38-53.

Klir, G.J. 1991. Facets of systems science. Plenum Press, New York, 664 pp.

Klir, J.G. and Bo Yuan. 1995. Fuzzy sets and fuzzy logic. Theory and applications. Prentice Hall, 574 pp.

Klir, G.J. and M.J. Wierman. 1999. Uncertainty-based information. Elements of generalized information theory. Physica-Verlag, 168 pp.

Kruse, R.,J. Gebhardt and F. Klawonn. 1994. Foundations of fuzzy systems. John Wiley & sons, 265 pp.

Landsberg, J.J. 1986. Physiological Ecology of Forest Production. Academic Press. London, 198 pp.

Landsberg, J.J. and R.H. Waring. 1997. A generalized model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. Forest Ecology and Management, 95:209-228.

McMurtrie, R.E. 1985. Forest productivity in relation to carbon partitioning and nutrient cycling: A mathematical model. In Attributes of trees as crop plants, Cannell, M.G.R. and J.E.Jackson (eds.) Inst. of Terrest. Ecol. Abbots Ripton, Huntington, U.K.

Mitchel, M. 1996. An introduction to genetic algorithms. Cambridge, Mass. MIT Press, 205 pp.

Morgan, M. G. and Henrion, M. 1990. Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis, Cambridge University Press, Cambridge, UK, 344 pp.

Neilson, R.P. 1995. A model for predicting continental scale vegetation distribution and water balance. Ecological Applications, 5:362-386.

Oderwald, R. G. and R. P. Hans. 1990. A methodology for constructing, estimating coefficients of, and testing process models. Forest Growth: Process Modelling of Responses to Environmental Stress, pp. 200-207.
Oderwald, R. G. and R. P. Hans. 1993. Corroborating models with model properties. Forest Ecology and Management, 62: 271-283.

Omlin, M. and P. Reichert. 1999. A comparison of techniques for the estimation of model uncertainty. Ecological Modelling, 115:45-59.

Parysow, P., Gertner, G. and Westervelt, J. 2000. Efficient approximation for building error budgets for process based models. Ecological Modelling, 135: 111-125.

Rasmussen, S. and C.L.Barret. 1995. Elements of a Theory of Simulation. Santa Fe Institute, working paper 95-04-040.

Ravetz, J. 1999a. Models as Metaphors, ULYSSES Working Paper WP-99-3, Darmstadt, University of Technology.

Ravetz, J. 1999b. Developing principles of good practice in integrated environmental assessments. International Journal of Environment and Pollution.11(3): 243-265.

Reynolds, J. H. and Ford, E.D. 1999. Multi-criteria assessment of ecological process models. Ecology, 80(2): 538-553.

Running, S.W., and J.C. Coughlan. 1988. A general model of forest ecosystem processes for regional applications. I Hydrological balance, canopy gas exchange and primary production processes. Ecological Modeling. 42:125-154.

Sangalli, A. 1998. The importance of being fuzzy and other insights from the border between math and computers. Princeton University Press, 173 pp.

Saltelli, A. 2000. What is Sensitivity Analysis? *In* Sensitivity Analysis, Edited by A. Saltelli, K. Chan and E.M.Scott. Willey Series in probability and statistics, pp 3-13.

Sommerville, I. 2001. Software Engineering. 6th Edition. Addison-Wesley, Usa, 693 pp.

Stroustrup, B. 1994. The C++ Programming Language. Addison-Wesley, Usa, 691 pp.

Sykes, M. T, C. Prentice and F. Laarif. 1999. Quantifying the impact of global climate change on potential natural vegetation. Climate Change, 41: 37-52.

Thornton, P.E. 1998. Description of a numerical simulation model for predicting the dynamics of energy, water, carbon and nitrogen in a terrestrial ecosystem. Ph.D. dissertation. University of Montana, Missoula, MT 280 pp. [Available from Mansfield Library, University of Montana Missoula, MT 59812].

Troelsen, A.W. 2000. Developer's workshop to COM and ATL 3.0. Wordware Publishing, Inc., Texas, 697 pp.

VEMAP Members. 1995. Vegetation/ecosystem modeling and analysis project: comparing biogeography and biogeochemistry models in a continental scale study of terrestrial ecosystem responses to climate change and CO2 doubling. Global Biogeochemical Cycles. 9(4):407-437.

VEMAP1. Kittel,T.G.F., N.A.Rosenbloom, T.H.Painter, D.S.Schimel, J.M.Melillo, Y.Pan, D.W.Kicklighter, A.D.McGuire, R.P.Neilson, J.Chaney, D.S.Ojima, R.McKeown, W.J.Parton, W.M.Pulliam, I.C.Prentice, A.Haxeltine, S.W.Running, L.L.Pierce, R.R.Nemani, E.R.Hunt, T.M.Smith, B.Rizzo, and F.I.Woodward. 1995. The VEMAP integrated database for modelling United States ecosystem/vegetation sensitivity to climate change. Journal of Biogeography, 22:857-862.

Weaver, W. 1948. Science and complexity. American Scientist. 36:536-544.

Zeigler, B.P. 1984. Theory of modelling and simulation. Krieger Publishing Company,Malabar, Florida, 435 pp.

Zimmermann, H.-J. 2000. An application oriented view of modeling uncertainty. European Journal of Operational Research, 122:190-198.