Regional-Scale Forest Production Modeling using Process-Based Models and GIS

A Plan B Paper
Submitted to the Faculty of the Graduate School
Of the University of Minnesota
By

Ryan William Kirk

In Partial Fulfillment of the Requirements
For the Degree of
Master of Science

____
Professor Thomas E. Burk
Adviser

____
Date

Acknowledgements

I wish to thank my adviser Dr. Thomas E. Burk for his guidance and patience during the prolonged process of working on this project. Thanks also to Dr. Paul Bolstad and Dr. Steve Manson for their assistance on various aspects of this project and involvement as examining committee members. I would also like to thank the public agencies that provided the data used in the study, including the Minnesota Department of Natural Resources, the USDA Natural Resources Conservation Service, the National Weather Service, and the US Forest Service.

I would also like to acknowledge my fellow students for initiating me to the realities of graduate school and instructing me on the processes and challenges of science, including Michael Counte, Melissa Arikian, Greg Edwards, Ryan Miller, Katherine Mitchell, Phil Radtke, Jason Henning, Jon Martin, Roy Rich, Mark Norris, Raju Vatsavai, Perry Nacionales, Chris Edgar, and Andy Jenks, among others. Lastly, I would like to send thanks to my family for their continued support during this change of career, and especially to my wife Jamie, who has inspired me in more ways than I ever thought possible.

Abstract

While research scientists have used process-based models of forest growth for several decades, forest managers have only recently begun to adopt them in production environments. This lag is accredited to the nature of process-based models, which are often difficult to parameterize, challenging to validate, and built around limited technical implementations. This project addresses these limitations by incorporating standard information system and Geographic Information System (GIS) concepts into the modeling framework.

As a sample implementation, the PnET-II and 3-PG models were run within a GIS for the Arrowhead region of northeastern Minnesota. The two models were each run at two spatial scales (1x1 km and 10x10 km grid cells). The mean NPP estimates across the two models and two scales range from 783.2 to 820.9 g C m^{-2} yr^{-1}, with standard deviations between 218.7 and 255.0 g C m-2 yr^{-1}. The two models produced similar predictions, and were comparable to published values from other studies in the region, including the USFS FIA database. The largest source of variation was forest cover type. Spatial aggregation of data sets had non-linear and non-uniform effects on the mean and variance of predicted NPP values. Predictions were most sensitive to changes in temperature values.

Based on the experiences of this modeling study and a review of the literature, a framework for implementing regional-scale process-based models within a GIS is presented. Primary components of the framework include ecological modeling considerations, data sources and stores, and technological processing requirements. Overall, two primary arguments are made in this project. First, technology can and should provide the link for continuing communication between model developers and forest managers. Second, in order for process-based models to be successfully incorporated in operational environments, improved information system designs are needed.

Table of Contents

List of Tables

Chapter 1

Chapter 2

List of Figures

Chapter 1

Chapter 2

Chapter 3

1 Introduction

The use of mathematical models in forest management has long been common practice. Traditionally, statistically-derived empirical models based on historical measurements are used to predict timber volume supplies and evaluate management practices for specific forest stands. Such models, while effective, are often limited in geographic scale and are inflexible to varying environmental conditions. Many modern forest management objectives are defined for large land areas such as watersheds or regions and account for changes in environmental conditions. Thus, these empirical models are not readily applicable to all management considerations.

Over the past three decades a new modeling paradigm has become popular in the scientific community. Rather than using empirical measurements, this “process-based” modeling approach attempts to simulate the general ecological mechanisms of a given ecosystem. With the remarkable advances in computing technologies and understanding of ecological processes, process-based simulation models are providing means to address scientific and management questions at all spatial scales, from individual trees to the entire globe.

Natural resource managers, however, have expressed concerns over using process-based models for practical decision-making at large spatial scales. One primary point of concern is the sheer complexity of implementing modeling studies of this magnitude. Significant sources of error may result if incorrect inputs, processing steps or interpretations are introduced. The project presented here addresses this concern. Specifically, a generalized framework for implementing regional-scale process-based forest productivity models is developed, based in part from an implementation and comparison of two widely used process-based forest simulation models.

1.1 Rationale

In the fall of 1998, members of the International Union of Forestry Research Organization (IUFRO) met in Saariselkä, Finland, to discuss the state of process-based modeling and the current application of process-based models to forest management. In the summary report (Mäkelä et al. 2000), the group affirmed the potential of process-based models as management tools across all spatial scales. A primary recommendation of the group was for improved practical implementations within operational management systems. In essence, a key factor to further use of process-based models as decision-making management tools is in information system design and analysis, not just continued scientific development.

Similarly, other summary publications have concluded that process-based models can be applied to management decision-making with improved implementations. Battaglia and Sands (1998) argue that current process-based models are overly complex for practical use and are in a state of constant development, and this is in disagreement with the desire for robustness and consistency in forest planning methodology (Sievänen and Burk 1993). Johnsen et al. (2001) contend that process-based models are quite valuable in simulating extremely complex forest systems, but will only be adopted when the complexities of research models are overcome. Korzukhin et al. (1996) conclude that with an increasing focus on ecosystem-based forest management, process-based models become a valuable tool for addressing a large variety of management decisions. Thus, with a diverse and clearly defined interest in process-based modeling for natural resource decision making, efforts need to be made to join the needs of forest managers with the powerful models being developed by researchers around the globe.

The scope of this project is to conduct and evaluate implementations of two process-based models applied at a regional scale. These two key components, process-based modeling and regional scale forest resource management are discussed individually.

1.1.1 Process-based forest growth modeling

Processed-based models can be defined as formalized statements of hypotheses regarding a complex system and its responses to stimuli (Landsberg 1986). To scientists, such models are tools that provide a structure for organizing current knowledge of a particular system, a framework with which to test hypotheses about that system, and a means to evaluate responses to stimuli within the system (Landsberg and Gower 1997). Due to the sheer complexity of calculations, process-based models are invariably presented as stand-alone computer programs or nested within spreadsheet applications. Thus, endless combinations of user interfaces, output, presentation, and analysis options are possible.

In recent years, forest managers have expressed interest in the application of process-based models in forest management decision making (Mäkelä et al. 2000, Korzukhin et al. 1996, Johnsen et al. 2001). In a summary paper, Battaglia and Sands (1998) identify five potential uses of process-based forest productivity models as management tools: (1) prediction of growth and yield, (2) selection of new plantation sites, (3) identification of site limitations on productivity, (4) assessment of risks associated with locations or management options, and (5) use of models as surrogates for field experiments. Mohren and Burkhart (1994) argue that process-based models provide greater potential for predicting forest growth under varying environmental conditions than empirical growth and yield models. From this perspective, the focus of modeling shifts away from scientific inquiry to strategic and operational considerations.

1.1.2 Modeling as a regional-scale management tool

Forest managers are interested not only in the characteristics of specific trees or forest stands, but also in trends that extend across large areas such as watersheds, landscapes or eco-regions (Shindler 1998, Gunderson et al. 1995). While scientifically based management at these coarse scales is desired, collecting appropriate data is a major challenge. For most organizations, collecting field data across large areas that meet even the most basic accuracy standards is logistically and financially impossible. Process models provide a viable alternative to large-scale field sampling for several reasons. First, the cost of data collection is largely reduced or removed completely. Second, advanced Geographic Information System (GIS) and remote sensing technologies provide data and analysis techniques unavailable with traditional sampling methods (e.g., He et al. 1998). Third, collecting data at a coarse-scale may increase accuracy and reduce errors compared to field measurements that have to be aggregated.

Although the benefits of process-models for regional scale analysis are quite evident, complexity remains a primary concern. It is important to investigate the possibilities of minimizing the complexity for end users in order for the benefits to be realized.

1.2 Objectives

The research presented here aims to provide support for bridging the gap between scientific and operational implementations of process-based forest production models. To meet this end, this project addresses four primary objectives:

1) Review the current state of regional-scale modeling research and application.

2) Implement a regional-scale modeling study using GIS and remote sensing technologies for northeastern Minnesota.

3) Compare model predictions of NPP at a regional-scale using two popular process-based models (PnET-II and 3-PG).

4) Based in part on the results of objectives 1, 2 and 3, develop a generalized framework for the technical implementation of regional-scale ecosystem process models.

The remainder of this chapter deals primarily with objective 1, providing a background and literature review, with specific emphasis on the models and modeling approach evaluated in this project. Chapter 2 addresses objectives 2 and 3. The third and final chapter of this report focuses on objective 4. A Portion of this project (particularly objectives 3 and 4 above) was condensed and published in the proceedings of the 4th Southern Forestry and Natural Resource GIS Conference (Kirk and Burk 2005), and is reproduced with this project in Appendix C.

1.3 Background and Literature Review

A remarkable surge of publications discussing the development and application of ecosystem process models in the past decade provides strong evidence for the level of interest in the models from both research scientists and, implicitly, their funding agencies. Several functional types of ecological process models have been developed, evaluated, and validated, from individual tree physiological models up to global carbon balance models. Ecosystem modeling for each of these functional types requires different types of considerations (for an in depth discussion of modeling considerations at varying ecological scales, see Meentemeyer 1989, Field and Ehleringer 1993, Waring and Running 1998).

This project addresses the implementation and use of one type of process-based model, commonly called ecosystem process models, at a single spatial scale generally referred to as a regional scale, with total area on the order of 10^2 to 10^5 km^2. Due in part to the technological advances in computing systems and remote sensing capabilities, regional-scale applications have been of particular interest as a new spatial level of study. The scope of this review is limited to regional-scale modeling and models that predict forest productivity. Other spatial scales and model types are considered only in relation to factors important to regional scale analysis.

1.3.1 Process-based ecosystem modeling

Ecosystem process models (EPMs) can be characterized by common traits found among different models. Typically, forest EPMs model the primary production portion of the terrestrial carbon (C) cycle (Figure 1). The C balance of forests is an important factor in the global C cycle; 80-90% of plant C and 30-40% of soil C are located in the forested areas around the world (Landsberg et al. 1995).

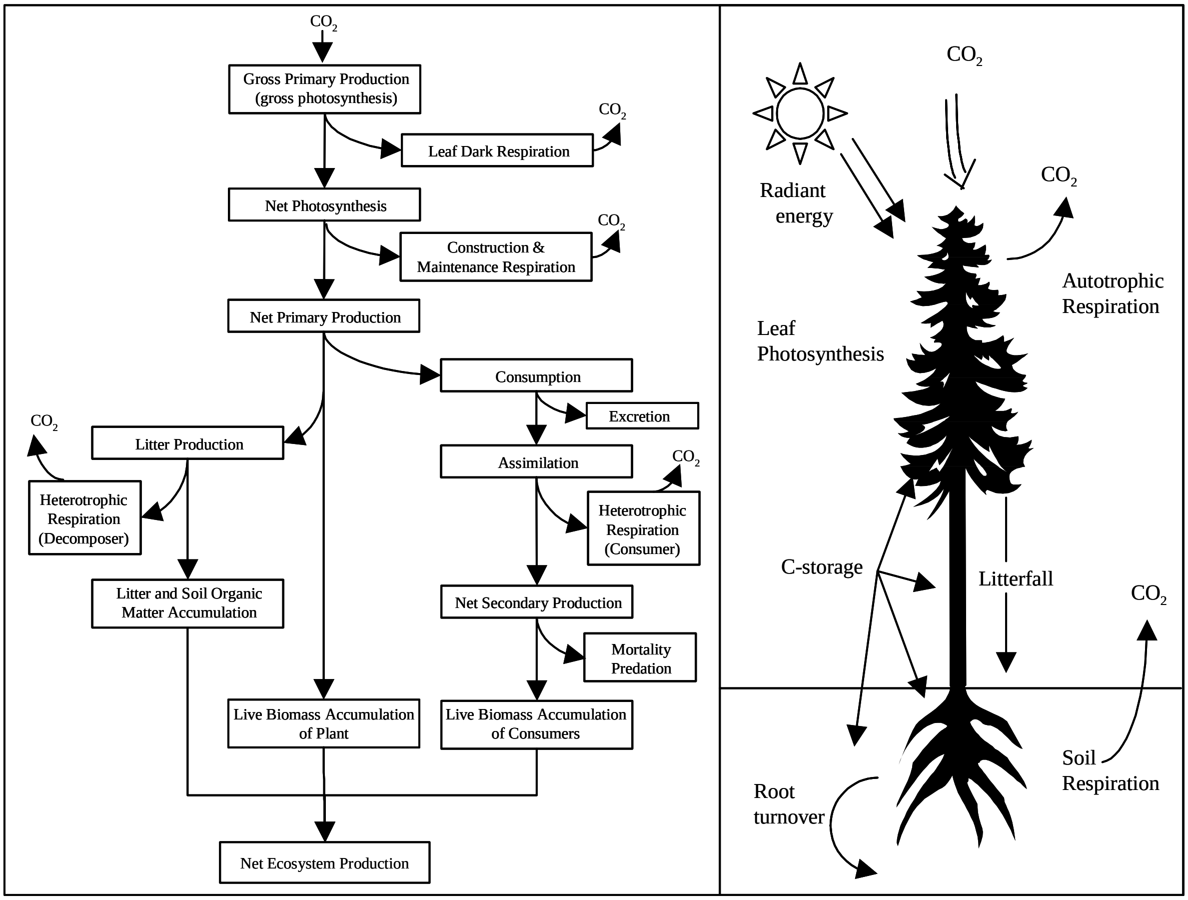


Figure 1: Diagram (left) and visualization (right) of the terrestrial carbon cycle. Adapted from Aber and Melillo (1991) and Landsberg and Gower (1997), respectively.

The primary output of forest EPMs is Net Primary Production (NPP). NPP is defined as the annual accumulation of organic matter (C) per unit of land for a given period of time (Schlesinger 1991). NPP is an important value for climate change research because it can be used as an indicator of the sequestration of atmospheric CO_2 by terrestrial ecosystems (Jiang et al. 1999). Forest managers are interested in NPP as a surrogate measure of volume growth of forest resources (Mäkelä et al. 2000).

Typically, EPMs that are applicable at regional scales model the C cycle by focusing on the processes of photosynthesis, respiration, and allocation of C within trees or a forest stand. The amount of photosynthesis and respiration is usually calculated with a “radiation-use efficiency” approach, in which total potential photosynthesis is determined and then reduced based on any number of environmental modifiers. Common environmental drivers are climatic factors, such as vapor pressure deficit, and water availability in the soil. The fact that C allocation routines vary greatly between models is indicative of the current lack of scientific understanding in C allocation processes (Landsberg and Gower 1997).

Regional-scale EPMs usually view a forest as a single homogenous unit (i.e., “Big Leaf” model) instead of as a set of individual trees. This allows predictions to have a scale-less dimension, meaning that the model can be run for any size forest stand or region. Other types of process-models can be classified as individual tree models (e.g., Robinson 1999), gap models (e.g., Pastor and Post 1986), and global carbon models (e.g., Raich et al. 1991). None of these other functional groups of process-based models are considered in this research.

In addition to internal structure, another trait of regional-scale EPMs is a heavy reliance on modern technologies. Complex computer algorithms are required to model all of the required processes. GIS and remote sensing technologies are regularly used to manage input and output data across large areas. Without these tools, the implementation of EPMs would not be possible.

The two models examined in this project, PnET-II and 3-PG, are generalized Big-Leaf type models applied at stand to regional scales. They were selected because of the widespread interest in them, their focus on generalized relationships and parameterizations, and the relatively few data input requirements. The two models are discussed in detail below, with focus on regional-scale application. Other models similar to these are also briefly discussed.

1.3.1.1 PnET-II model

The original PnET (Photosynthesis and Evapotranspiration) model (Aber and Federer 1992) was developed as a generalized ecophysiological model of forest water and carbon balances. The model departs from previously published models by specifically attempting to condense complex physiological attributes of tree species into a few, lumped generalized processes (Landsberg and Gower 1997). A primary objective of the original research was to provide a simple, well-validated model of forest ecosystem production for application to issues of environmental change (Aber et al 1993). Several versions of PnET have been developed over the past decade (Table 1). Of these, PnET-II is the most widely used in the PnET family of models and is currently the only version used in studies involving regional-scale applications. Thus, PnET-II is the only model in the PnET family considered in this project.

Table 1: Popular versions of the PnET family of ecosystem process models.

PnET Version	Source	Description
PnET	Aber and Federer (1992)	Original model
PnET-II	Aber et al. (1995)	Enhanced version of PnET
PnET-DAY	Aber et al. (1996)	Daily time-step variation
PnET-CN	Aber et al. (1997b)	Extended version that models nitrogen cycle as well as carbon and water cycles
PnET-BGC	Gbondo-Tugbawa et al. (2002)	Expanded biogeochemical model

PnET-II is a lumped-parameter model of carbon and water balances that combines process-based and empirical components (Figure 2). PnET-II runs on a monthly-time step and has no specific spatial dimension, although it is commonly applied at a small watershed to regional scales (Aber et al 1995). Although PnET-II was designed for forest ecosystems, it has been calibrated for other vegetation types (e.g., Reich et al 1999). The model runs to a steady state with no consideration of mortality or succession.

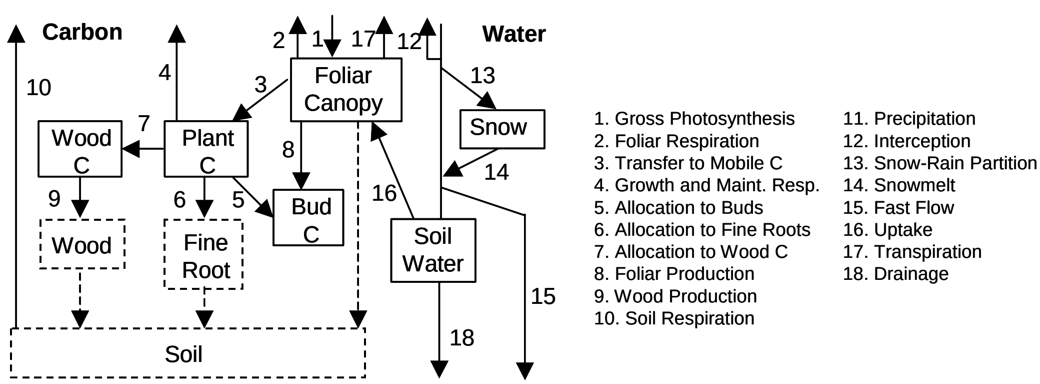


Figure 2: Structure and processes of PnET-II (from Aber et al. 1995).

Total canopy photosynthesis is calculated by integrating photosynthesis across canopy layers within a “green sponge” canopy architecture, where the canopy is represented as a single homogenous mass of thickness equal to the Leaf Area Index (LAI, unit area leaves per unit area ground) (Landsberg and Gower 1997). After subtracting out canopy respiration, the maximum net C gain is reduced by any limitations of three environmental drivers: temperature, vapor pressure deficit, and soil water deficit. Net gains in C are accumulated throughout the year in one pool and allocated at year-end to foliage, wood, and fine root pools. Carbon is allocated first to build leaf area, calculated as a function of growing degree days and respiration, and then to build root area, calculated as a function of leaf area. Wood carbon is calculated as the remaining C after leaf and root C are allocated.

Standard input variables for PnET-II are categorized as site variables, climate variables and vegetation parameters. Site variables include vegetation type, latitude, soil water holding capacity and initial biomass content. Climate variables include average monthly values for minimum and maximum daily temperature, daily precipitation, and daily amount of incoming photosynthetically absorbed radiation (PAR, the portion of the electromagnetic spectrum from which plants draw energy for photosynthesis). Vegetation parameters, required for each vegetation type, consist of quantitative values that define how a given vegetation type responds to environmental conditions. For example, different vegetation types will have varying optimal growing temperatures for photosynthesis. Specific PnET-II parameter values are discussed in Chapter 2, and parameter values for various forest types are presented in Appendix A.

Outputs for PnET-II are provided annually for both the carbon and water cycles. Vegetation growth outputs include gross and net photosynthesis, foliage, root, and stem wood biomass, and net ecosystem production (NEP, defined as gross photosynthesis minus respiration). Hydrologic outputs include annual precipitation, evapotranspiration, and drainage.

PnET-II has been used for several regional scale research applications (Table 2). These studies can be classified into three groups: ecophysiological studies (Aber and Driscoll 1995, Jenkins et al. 1999), hydrologic studies (McNulty et al. 1994, Aber et al. 1995, Bishop et al. 1998), and large area forest production estimation studies (Aber et al. 1995, McNulty et al. 1996, McNulty et al. 1997, Goodale et al. 1998, Ollinger et al. 1998, Mickler et al. 2002). One exception is a recent article (McNulty et al 2000), which uses PnET-II connected to a biogeography model and an economic model to examine the effects of climate change on forest growth and timber markets. In these studies, the spatial scale (grid cell size) ranges from 1 km^2 to greater than 50 km^2 (Table 2).

Without exception, the authors in these studies concluded that PnET-II provided a useful, viable tool for addressing the varying regional-scale objectives. Weaknesses of PnET-II for various applications were also identified. McNulty et al (2000) noted that PnET-II does not consider disturbance (e.g., forest thinning, fire, or herbivory), which can limit its application to forest management. McNulty et al (2000) mentioned that relying on limited data for parameters (most notably foliar N) are a key concern for model projections. Jenkins et al. (1999) suggested that model assumptions caused differences in PnET-II predictions compared to another model, TEM (Raich et al 1991). Thus, a solid data set for inputs and parameters that drive the model, such as foliar N and soil water holding capacity, is vital. Similarly, Bishop et al. (1998) concluded that generalized PnET-II predictions underestimated regional-scale hydrologic response, indicating that generalization may lead to bias for some, but not all, outputs. These weaknesses are similar to the general weaknesses of process-based modeling previously discussed.

Table 2: Summary of regional scale applications using PnET-II. Grid Cell Size is the spatial resolution of each cell for which the model is run.

Source	Description of Research	Grid Cell Size
McNulty et al. (1994)	Evaluated regional hydrologic response of loblolly pine to climate change in the southern U.S.	0.5° x 0.5° ~ 50 km x 75 km
Aber et al. (1995)	Predicted effects of climate change on forest production and water yield in the northeastern U.S.	1 km^2
McNulty et al. (1996)	Examined role of climate change on regional loblolly pine production in the southern U.S.	0.5° x 0.5° ~ 50 km x 75 km
Aber and Driscoll (1997)	Evaluated land use, climate variation and N deposition on forest N and C cycles in northern U.S.	0.5° x 0.5° ~ 40-50 km^2
McNulty et al. (1997)	Explored input scale effects on forest productivity and hydrologic yield in the southern U.S.	0.5° x 0.5° ~ 50 km x 75 km
Bishop et al. (1998)	Compared estimates of long-term hydrologic runoff in the northeaster U.S.	1' x 1' ~ 1.8 km^2
Goodale et al. (1998)	Predicted sensitivity of forest production to site quality and climate change in Ireland.	1 km^2
Ollinger et al. (1998)	Estimated regional patterns of forest production in the northeastern U.S.	30" x 30" ~ 1 km^2
Jenkins et al. (1999)	Identified sources of variability in forest production estimates in the northeastern U.S.	0.5° x 0.5° ~ 40 km x 50 km and 60" x 60 " ~2 km^2
McNulty et al. (2000)	Explored linkages between forest growth, biogeography and economic models in the southern U.S.	0.5° x 0.5° ~ 40 km x 50 km
Mickler et al. (2002)	Predicted current and future forest biomass levels to assess forest productivity and wildland fire fuel potential.	0.5° x 0.5° ~ 40 km x 50 km

1.3.1.2 3-PG model

The 3-PG (Physiological Processes Predicting Growth) model is a relatively new ecosystem model receiving a lot of attention in the forest modeling community. Developed by Landsberg and Waring (1997), 3-PG was designed for use as a practical forestry tool as well as a research tool. As with PnET-II, 3-PG is a generalized, monthly time-step radiation use efficiency model of forest growth that includes both process-based and empirical components. A key difference with 3-PG is the inclusion of allometric relationships, which means that biomass outputs are converted into units important to forest managers (e.g., stand density and volume).

Simplified relationships among several key ecophysiological processes characterize 3-PG (Figure 3).

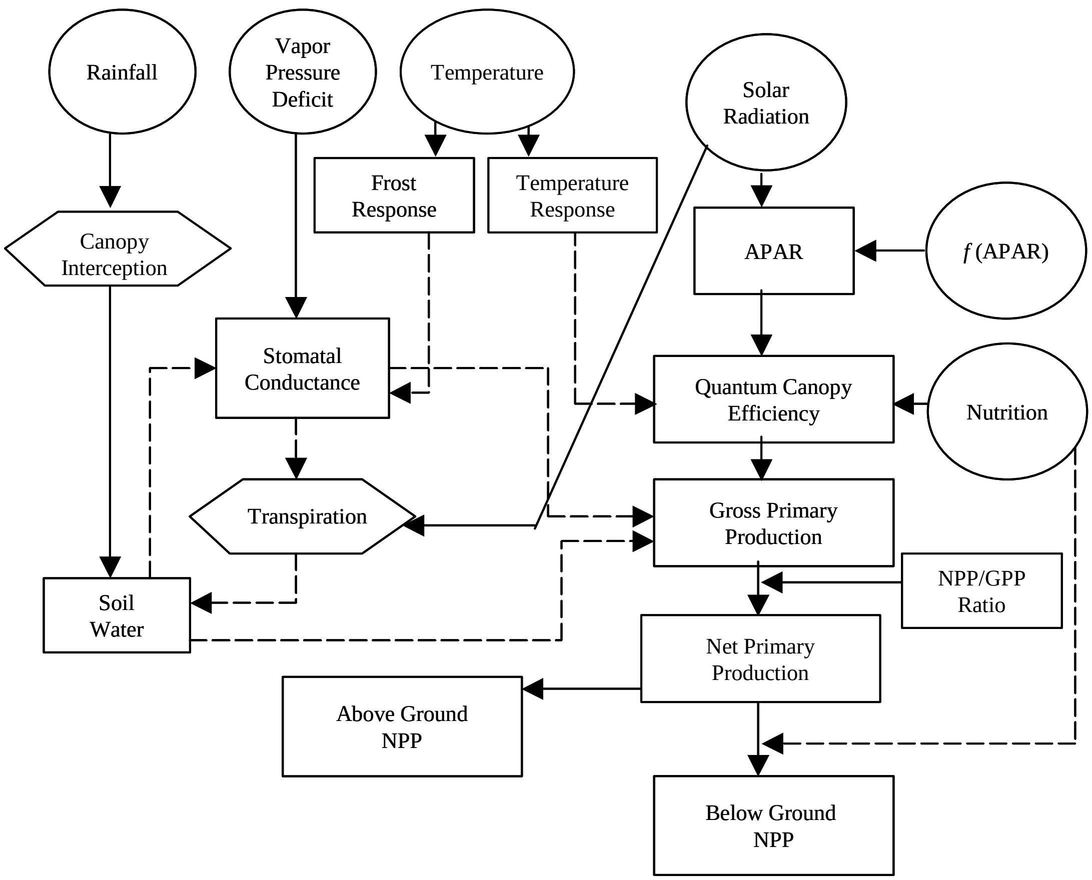


Figure 3: Processes of the 3-PG model (from Tickle et al. 2001).

Potential photosynthesis is calculated from the amount of absorbed photosynthetically active radiation (APAR) and reduced for any of four environmental modifiers: frost, soil water deficit, vapor pressure deficit, and age. The age modifier is unique to 3-PG, and follows the observation that as forests age, productivity decreases (Waring and Schlesinger 1985). GPP is determined as a set percentage of APAR as defined by a constant quantum canopy efficiency (i.e., the initial slope of the light response curve for a given species). NPP, in turn, is calculated as a fixed fraction of GPP. This species-specific NPP/GPP fraction (Waring et al. 1998) greatly simplifies model complexity, as respiration calculations are no longer required. C is allocated first to roots relative to the harshness of the environment (as defined by the environmental modifiers), and then to stem and foliage using a simple stem:foliage allometric equation.

3-PG site and climate inputs, for the most part, are similar to those for PnET-II. Required climatic variables are solar radiation, precipitation, and temperature. Site values include initial biomass, available soil water, and current stand age and density. Vegetation parameters include quantitative values describing physiological as well as allometric relationships. Specific parameters are not discussed in this review chapter. Key 3-PG parameters are discussed in Chapter 2. Parameters for several species and cover types are listed in Appendix A.

Annual outputs for 3-PG include NPP and root, stem, and foliage biomass. Unlike PnET-II and most other regional-scale EPMs, however, 3-PG goes beyond calculated NPP and converts outputs to variables important to forest resource managers. With number of trees per acre as an input value, 3-PG converts NPP into stand volume, stand density, average stem diameter, and mean annual increment (MAI, average annual stand growth) using general allometric equations. These statistics are vital information for most traditional forest management decisions.

In light of how new 3-PG is, few regional scale studies have been conducted using the model (Table 3).

Table 3: Summary of regional scale applications using 3-PG. Grid cell size is the spatial resolution of each cell for which the model is run.

Source	Description of Research	Grid Cell Size
Tickle et al. (2001)	Predicted forest productivity over a diverse 50,000 ha Eucalypt forest in Australia.	25 m^2
White et al. (2000)	Estimated forest and scrub biomass accumulation in New Zealand.	1 km^2
Coops et al. (1998)	Assessed use of satellite-derived estimates of forest photosynthetic capacity for user with 3-PG in Australia.	8km^2
Coops et al. (2001)	Compared patterns of NPP and water use using 3-PG and BIOME-BGC in southwestern Oregon, U.S.	1 km^2
Coops and Waring (2001a)	Estimated forest productivity using satellite-derived inputs in Oregon, U.S.	1 km^2
Coops and Waring (2001b)	Assessed forest growth under various climate change conditions in Oregon, U.S.	1 km^2
Coops and Waring (2001c)	Explored the use of multiscale remote sensing imagery to drive estimates of forest growth in Oregon, U.S.	1 km^2

As with PnET-II research, 3-PG was found to be a useful tool for regional scale forest growth modeling in all documented studies. These large area studies can be separated into two groups: ecophysiological studies (White et al. 2000, Coops et al. 2001, Coops and Waring 2001a) and forest production studies (Tickle et al. 2001, Coops et al. 1998, Coops and Waring 2001b). 3-PG has been shown to predict forest stand growth estimates for a large variety of forest types (Table 4), although at this time most research has focused on conifer and Eucalypt species. In these studies, 3-PG predictions have been correlated with field-based forest measurements of traditional statistics such as stand volume and density. In addition, there is presently one documented case of 3-PG being used in an operational environment on a Eucalyptus plantation in South America (Almeida et al. 2002).

Table 4: Species/forest types used in 3-PG studies. For these forest types, 3-PG stand volume, density and growth estimates have been correlated with field measurements.

Species/ Forest Type	Location	Source
Shasta Red Fir	Oregon, U.S.	Coops et al. 2001
Ponderosa Pine	Oregon, U.S.	Coops et al. 2001
Douglas Fir	Oregon, U.S.	Coops et al. 2001
Sitka Spruce	Great Britain	Waring 2000
Eucalypt spp.	Australia	Tickle et al. 2001
Patula Pine	South Africa	Dye 2001
Radiata Pine	Australia	Landsberg et al. 2002
Norway Spruce	Sweden	Landsberg et al. 2002
Jack Pine	Ontario, Canada	Peng et al. 2002
Loblolly Pine	North Carolina, U.S.	Landsberg et al. 2001
Hard Beech	New Zealand	White et al. 2000
Mountain Beech	New Zealand	White et al. 2000

1.3.1.3 Other regional-scale ecosystem process models

While PnET-II and 3-PG are popular Big-Leaf EPMs, they are not the only alternatives. Several other models, although not considered in this research, deserve mention. FOREST-BGC (Running and Coughlan 1988, Running and Gower 1991) is a widely used EPM for modeling C, N, and water cycles in forested ecosystems. Requiring only an estimate of LAI to represent canopy architecture allows FOREST-BGC to be usable for regional-scale analysis. FOREST-BGC has been used in a wide variety of research applications (e.g., White and Running 1994, Running et al. 1989) and is currently part of NASA's Earth Observing System (EOS) MODIS Land Science program. The CENTURY model (Parton et al. 1987, Parton et al. 1993) is a general model of water, C, N, and selected other nutrient cycles in grassland and forest ecosystems. With a focus on soil processes, CENTURY provides a different perspective on calculating NPP and is used primarily in research. TEM (Raich et al. 1991) models C and N cycles across large land areas (i.e., continent to global scale) at a 0.5 degree latitude and longitude resolution. While TEM predictions are correlated with finer scale predictions (Jenkins et al. 1999), the coarse resolution limits its applicability for forest management decision making.

Each of these alternative models, as well as several others, have received a lot of attention and secured their place in the ecological research community (e.g., VEMAP members 1995). Although this research focuses only on PnET-II and 3-PG, reviewing research that involved these other models was valuable. Future research on process-based forest production models for operational decision-making should consider the strengths and weaknesses of a wide variety of models and their implementations.

1.3.2 Regional-scale natural resource analysis

Producing clear definitions and boundaries of what constitutes a specific ecological scale is a difficult task. Ecological scales (Figure 4) are identified based on hierarchy theory, whereby a complex system is comprised of inter-related subsystems in a hierarchical fashion (Simon 1962).

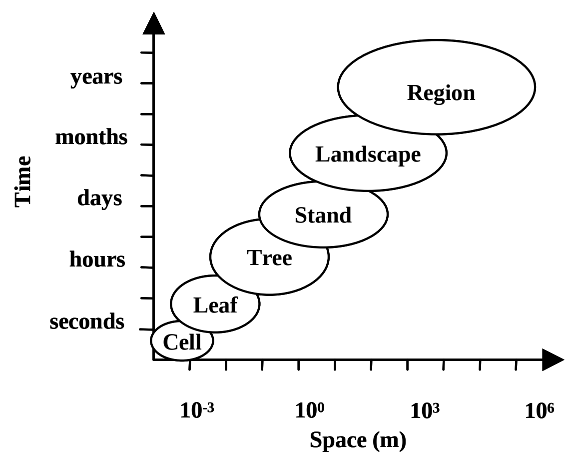


Figure 4: Ecological scales as related to temporal and spatial scale (adapted from Jarvis 1995).

This research focuses entirely on regional scale modeling. Regions, in this context, can be defined as broad geographical areas with a common macroclimate and sphere of human activity and interest (Isard 1975). Thus, regions have both ecological and cultural components, and can be distinguished by relatively similar patterns of climate, geologic history, nutrient supplies and land use history.

Forest managers regularly make decisions across large land areas as well as in local forest stands. These decisions often fall along political (e.g., county or state) or natural boundaries (e.g., watersheds or ecoregions). Management decisions at this scale are often holistic, requiring a balance of commodity and ecological objectives (Baskent and Yolasigmaz 1999).

With an emphasis on spatial patterns and ecological processes, regional scale process-based modeling provides a valuable tool for addressing these complex and diverse objectives. For example, Mickler et al. (2001) concluded that regional scale forest growth modeling studies could be used by forest managers to examine the role of climate change on forest production (NPP), fuel loads and fire regimes.

1.3.3 Technologies for regional scale natural resource modeling

Regional-scale process based modeling would likely be impossible without the use of two related technologies: Geographic Information Systems and remote sensing. EPMs by nature are extremely complex, requiring a large and diverse data set. By incorporating spatial data from a variety of sources, including satellite imagery, variations in environmental conditions across large areas can be more effectively included in forest production modeling.

Geographic Information Systems (GIS) are defined as computer-based systems that aid in the collection, maintenance, storage, analysis, output and distribution of spatial data and information (Bolstad 2002). EPMs benefit from GIS for several reasons. First, required input data can be acquired from a wide variety of sources and organized seamlessly. Second, GIS provide a spatially explicit structure from which to run models, even if the models themselves are not spatially explicit. A model can be run for each cell in a raster grid and the results displayed in map form. Third, advances in the science of GIS have led to a new suite of spatial analysis tools that can be used to analyze patterns of model inputs and outputs. Currently, two versions of the 3-PG model have been designed around a GIS. 3-PGS (Coops et al. 1998) uses satellite-derived estimates of photosynthetic capacity in a grid cell format as a primary input. 3-PGSPATIAL was designed to include spatial layers of climate and soils data as direct inputs. Although PnET-II does not have a spatially explicit version, it is linked to a GIS relatively easily (Ollinger et al. 1998).

A second key technology for regional scale modeling is remote sensing. In this context, remote sensing can be defined as observation of the earth's surface by means of reflected or emitted electromagnetic energy (Campbell 1996). Remote sensing data are typically collected by airborne or satellite sensors and often consist of a stack of layers containing information about different segments of the electromagnetic spectrum (e.g., green, red, or near infrared bands). Remote sensing data are often classified to provide land cover/land use data. These land cover/land use maps are frequently used as the basis for regional scale forest growth modeling (e.g., Ollinger et al. 1999, Coops et al. 1998). In addition to cover type, remote sensing technology provides means to estimate important biophysical attributes, such as LAI, across large land areas (e.g., Gower et al. 1999). Obtaining these data by traditional ground sampling methods would be logistically and financially unfeasible for all but the biggest budgets. Together, remote sensing and GIS technologies provide the means as well as the tool set for effectively modeling regional scale forest growth.

1.4 Data and Tools

Regional scale EPM studies typically rely on a combination of publicly available and locally developed data. The most important data requirement is land cover/land use for which the models are parameterized. Cover type maps are usually interpreted from satellite data (e.g., Ollinger et al. 1999, Running et al. 1989), extrapolated from nationwide forest inventory data (McNulty et al. 1994), or developed locally. Climate data are either interpolated from regional climate stations or from climate models (Hungerford et al. 1989, Ollinger et al. 1995). Soils data are estimated from local research or from the national STATSGO soils database (SCS 1991). Forest stand data are summarized from local estimates or from state or nationwide forest inventory data. Parameter values are estimated from local studies or comprehensive literature searches. This diverse range of data sources implies that no data are ideal for all studies, and that specific care should be taken to obtain the most accurate and precise data possible.

This project is centered on the implementation of two regional scale EPMs for the Arrowhead region of northeastern Minnesota. In order to assess the robustness of the models, only freely available and widely used data sets were considered and included. All processing was completed using popular software and programming packages, including ESRI ArcView 3.x, ArcInfo and ArcGIS 8.x GIS software, ERDAS Imagine 8.x image processing software, the C programming language, and the Perl 5.x scripting language. Chapter 2 discusses the specific input values and processing steps in detail.

1.5 General Methodology

Methodologies used in regional scale modeling studies vary greatly based on resources available and study objectives. However, three distinct phases can usually be identified in different studies. The first phase involves collecting and compiling the input data and parameters. This phase involves any field data collection, analysis, and GIS processing to prepare each unit for which the model will be run. This often requires the reorganization of GIS data into a format suitable for model runs. The second stage consists of running the models and performing sensitivity analyses on model predictions. Sensitivity analysis consists of repeated model runs where data input values are systematically changed for the purpose of assessing how sensitive model predictions are to varying inputs. The third and final phase involves validating the model predictions against field measurements or alternative estimates. Sensitivity analysis and validation help identify sources of bias or error and can help to quantify confidence in model predictions.

The methodologies used in this research are limited to popular and well-documented options. Selected methodologies are described in detail in chapter 2. Chapter 3, which describes a framework for implementing regional scale modeling studies, contains descriptions of widely accepted alternative methodologies to those used in this project.

2 Regional Scale forest production modeling in northeastern Minnesota

2.1 Introduction

Ecosystem process models (EPMs) have several potential applications. To date, these applications have generally been academic, such as testing hypotheses regarding ecosystem processes or evaluating the effects of global change on forest growth. However, there is a large amount of interest in the application of EPMs in predicting forest production over large land areas.

This chapter presents the results of a regional-scale modeling study within a GIS for the Arrowhead region of northeastern Minnesota. The objectives of this modeling project are to:

1) estimate forest biomass growth for the Arrowhead region of northeastern Minnesota,

2) compare model predictions for two popular models at two spatial scales, and

3) perform a sensitivity analysis to evaluate model inputs.

The overarching objective of this research is to evaluate the process of implementing regional scale EPMs. Chapter 1 provided background and rationale. The objective of the final chapter, Chapter 3, is to identify the core components of regional scale modeling and develop a technical and conceptual framework for implementing such models. The experiences gained in this modeling project (Chapter 2) provide much of the basis for the modeling framework. While this chapter focuses on the methods and results of the model predictions, critical evaluation of the modeling process is presented in Chapter 3.

2.2 Study Area

The study area for this project is a 5 county area in northeastern Minnesota, an area commonly called the “Arrowhead” region of Minnesota (Figure 1). Receding glaciers in the last

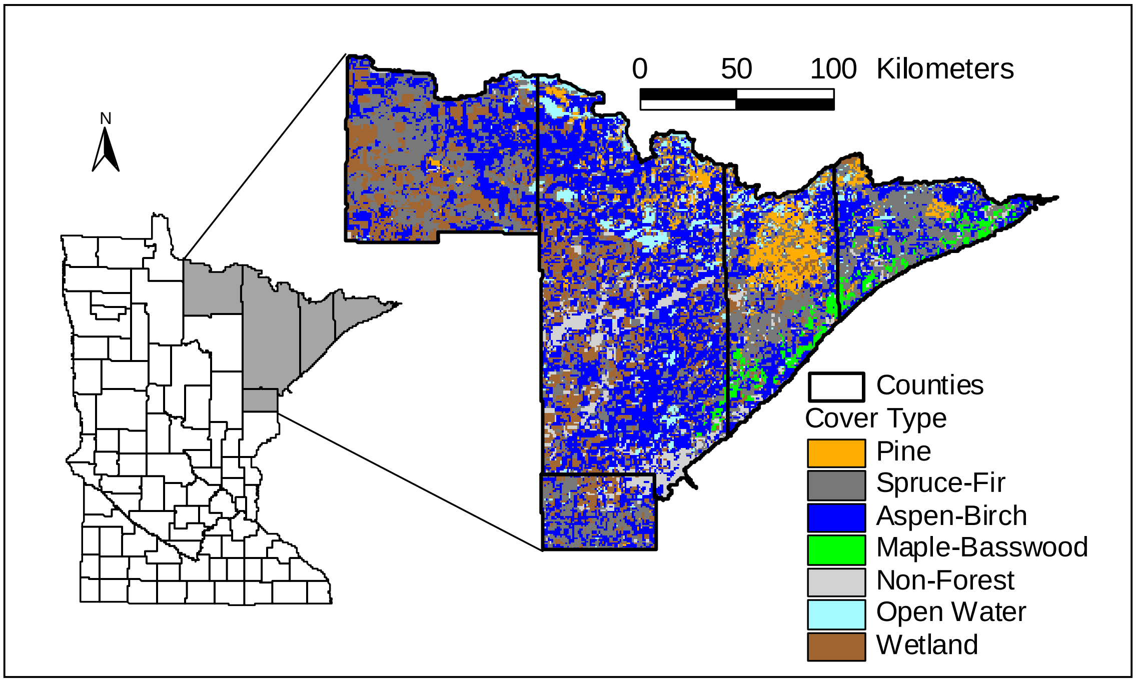


Figure 1: Study area in northeastern Minnesota.

ice age (less than 10,000 years ago) shaped the region, which is geologically identified as the Superior Uplands portion of the Canadian Shield. As a result, the area is characterized by exposed bedrock, extensive mineral deposits, varied topography, thin soils, and a large abundance of lakes. The area is bordered on the east by Lake Superior and on the north by Canada. The climate is characterized by cold winters, cool summers and an average of 66 cm of precipitation per year. In terms of vegetation, the region falls in a transition zone between temperate and boreal forests, containing predominantly aspen, birch, and white, red, and jack pine in the uplands, and spruce, fir and tamarack in the lowlands.

2.3 Data

Both spatial and aspatial data were required for this project. At the onset, we decided to use generalized and publicly available data sets. Although this limitation was in small part due to budget constraints, the primary reason was to evaluate model predictions based on a lower-limit of data quality. In other words, we wanted to evaluate how well the models worked using the most generic data available.

All of the data, with the exception of model parameters, were organized within a GIS. Data collection and development were the primary emphases for the early phases of this project. The required data can be grouped into four general categories: forest inventory data, climate variables, site variables, and model parameters. Each of these data categories, as well as the collection and development processes, is described below. Sample maps of key variables are included in Appendix A.

2.3.1 Forest Inventory and Analysis (FIA)

The primary mission of the U.S. Forest Service FIA unit is to quantify and monitor forest conditions across the entire country (Jakes, 1980; Leatherberry et al., 1995). Through extensive, periodic surveys, the FIA group provides forest area, volume, growth, disturbance, and mortality estimates. These data are used for a vast variety of research and applied applications.

The time period considered in this model consists of the years between the 4^{th} and 5^{th} Minnesota FIA surveys, 1977-1990 (Jakes, 1980; Leatherberry et al., 1995). The study area follows the boundary of the Minnesota Aspen-Birch FIA unit, which encompasses 5 counties in northeastern Minnesota (Figure 1). Also, the four forest types evaluated in this analysis were aggregated from the general FIA forest type classes. In essence, the FIA surveys provided the scope of this project.

Two types of FIA data are used in this project. First, the forest cover type map used in this analysis was derived from the 1977 Minnesota FIA survey map (Jakes, 1980). The FIA survey uses a stratified sampling inventory design based on cover type. Prior to any field visits, forest cover types were delineated off of aerial photographs and summarized in a statewide map.

Paper copies of the cover type map were delivered with the 1977 summary report. No digital copies are known to exist. As a result, the paper map was converted to a digital image with a four-step process. First, the paper map was scanned on a Microtek 9600XL flatbed scanner. Second, the map was georegistered to a geographic coordinate system, matching the lines of longitude and latitude included on the map. Third, the image was classified based on the map's cover type color schema. The final step involved converting the image to 1x1 km resolution raster grid. A second, 10x10 km resolution raster grid was created from the 1x1 km grid using a majority-area decision criterion.

This classification and aggregation process undoubtedly introduced errors into the cover type data. For cells containing multiple cover types or distortions (e.g. highway lines or paper creases) the cover type was manually classified. Approximately 4000 of the over 25,000 grid cells in the study area required manual classification. Similarly, when the majority-area aggregation failed (i.e., when two classes had the same total area within a cell) the class was assigned manually. As a quality control measure, several of the manually classified cells where changed in order to maintain the proportions of each cover type described in the FIA tabular summaries. I feel that the errors resulting from this process are acceptable since we are interested in the overall trends, not in predictions for specific areas.

The second type of FIA data used in this project are forest volume and growth estimates used to validate model predictions. These standard data are provided by the FIA Eastwide Data Retrieval database (Hansen et al., 1992) and have been used in other forest production modeling studies (Brown and Schroeder, 1999). The FIA data were summarized by general forest type, and are used primarily for model validation.

2.3.2 Climate variables

Both 3-PG and PnET-II require monthly temperature, precipitation and solar radiation values as inputs. Long-term monthly temperature and precipitation norms were acquired for several climate stations within and near the study area (Baker et al. 1985; Minnesota Climatology Working Group, 2003), and linked to the USGS Geographic Names Information System spatial locations (U.S. Department of Interior, 1987). A series of raster grids were interpolated from these climate station data, including monthly average rainfall and monthly minimum, maximum and mean temperature. The interpolation method for all layers consisted of an Inverse Distance-Weighted (IDW) algorithm with squared exponential weighting. Thiessen Polygon interpolation and several different IDW weightings were considered. However, a squared exponential IDW approach appeared to most smoothly depict trends across all variables. Since different interpolation methods can provide a range of predictions, it is recommended that additional interpolation methods be considered in any further research.

Freely available solar radiation data from the National Solar Radiation Database were found for only two cities within the study area, Duluth and International Falls (Maxwell et al., 1995). For these data, monthly grids were linearly interpolated along a latitudinal gradient, which produced only a minor change in solar radiation across the entire study area. This is not considered to be a limitation, particularly considering the moderate spatial extent of the study and the coarse spatial resolution of the model runs. Sample maps of key climate variables are included in Appendix A.

2.3.3 Site variables

The primary site data sets required for the two models are forest type and soil variables. The forest type (also called a cover type) is the general vegetation type for which the two models are parameterized. For this study, four general forest types are used: Aspen-Birch (39% of the study area), Spruce-Fir (26%), Pine (5%), and Maple-Basswood (2%). The remainder of the study area (28%) consists of non-forest areas, including open water, wetlands, developed areas, and bare land. The models were not run for these non-forest areas. Forest type data were adapted from the 1977 FIA forest type map as described above in section 2.3.1. Although the four forest types are not the only forest types found in the study area, they represent the dominant.

The two soil variables of interest were soil water holding capacity (SWHC) and soil texture type. SWHC is the total water capacity available to plants within the soil horizon. SWHC data were derived from the nation-wide STATSGO soil data set (SCS, 1991) using the following calculation:

$$SWHC = \sum \frac{\text{awcl} + \text{awch}}{2} * (\text{laydepl - laydeph}) \qquad [1]$$

where *awcl* and *awch* are the available water capacity minimum and maximum for a given soil layer, respectively, *laydepl* is the depth from the soil surface to the upper boundary of the layer, and *laydeph* is the depth from the soil surface to the lower boundary of the layer. These variable names are standard for the STATSGO database. Lathrop et al. (1995) found the STATSGO and SSURGO databases provided unreliable estimates of SWHC for the New England area. They elected to use a constant SWHC value for all areas. We feel that the large variation of SWHC in the northeastern Minnesota is a significant factor in determining forest growth. Thus, we accept the errors in the data and feel that the overall trends are more important than errors at specific sites.

In addition to forest type and SWHC, 3-PG requires an additional soil variable and several forest stand variables. The soil variable is soil texture class, which is also available in the STATSGO database. Processing soil texture class consisted of reclassifying the STATSGO texture classes into the 3-PG texture classes. Required forest stand characteristics include initial standing biomass, stand density (trees per hectare), and stand age. For initial model runs, these values were extracted from distributions in the FIA database for each cover type and randomly assigned to grid cells. However, this significantly altered production estimates. 3-PG applies an age modifier in the primary production routine (see section 1.3.1.2 in Chapter 1), which lowers production for older forest stands. In addition, 3-PG includes a mortality component in the form of the "-3/2 self-thinning rule". By randomly assigning both density and age classes, unrealistic combinations of two values were included, which led to a huge variation in production estimates. Although the age and mortality modifiers are important to forest production estimates, they appear to require a more detailed and spatially explicit data set than was compiled for this project. As a result, a standard set of starting conditions was used for each forest type in order to remove these large effects. With this decision, average forest production predictions from both models could be compared.

2.3.4 Parameters

Process-based ecosystem models use parameter values to distinguish the characteristics and responses to stimuli of different vegetation types. In this project, four general forest types were selected: Pine, Aspen-Birch, Spruce-Fir, and Maple-Basswood. A parameter file, consisting of several dozen parameter values, is required for each forest type and each model. The compilation of parameter values for the two models is discussed below.

For PnET-II, parameter files for the Aspen-Birch, Maple-Basswood and Pine types were previously developed for a modeling project at the Cedar Creek Natural History Area in central Minnesota (Reich et al., 1999). For the Spruce-Fir type, all parameter values were the same as those of the BOREAS research project in Canada (Sellers et al., 1997).

For 3-PG, no known complete parameter sets were found for northern temperate hardwood forests. Documented parameters exist for conifer forests of the Pacific Northwest (Coops and Waring, 2001a), loblolly pine in the southern states (Landsberg 2001), and a variety of forest types in Australia and New Zealand (Landsberg and Waring, 1997). Thus, parameter sets had to be estimated for the four forest types used in this study. Where available, individual parameter values were found in the literature. For the remaining parameters, either the default values or estimates based on other documented parameters were used. A series of test scenarios were run using the range of estimated values in order to ensure the parameter combinations were comparable to the PnET-II predictions and relatively near documented NPP estimates in the area (e.g., Reich et al, 1999, Goetz and Prince, 1998). Thus, at least in part, the 3-PG parameters were calibrated against the PnET-II parameters. This may provide a bias in modeled estimates. In further research, these parameters should continue to be assessed and refined.

2.4 Methods

The overarching theme of this research is evaluating the technical implementation of EPMs. As a result, the description of methods emphasizes issues related to model implementation. This means that, in effect, the models are being treated as “black boxes.” Four methodological categories are discussed: data preparation, technical implementation, sensitivity analysis, and spatial analysis.

2.4.1 Data Preparation

All of the spatial data were organized within a GIS. Two primary lattice grids (in the form of point feature shapefiles) were created for the study area, one for each of the spatial scales of analysis (1x1 km and 10x10 km cells). Each lattice point represents the center location of a grid cell at the given spatial resolution. The attribute data for all required environmental and site data (described above) were spatially joined with these primary shapefiles, resulting in shapefiles with over 100 attributes. This attribute table was then extracted to a comma separated value (csv) file that was used for model runs.

2.4.2 Model Implementation

Both 3-PG and PnET-II have multiple model formats, including some that are spatially-explicit. For this study, we elected to use comparable versions that consisted of stand-alone software executables compiled from C/C++ source code. The 3-PG "Console" version is available from the developers at no cost[1]. PnET-II is also freely available in a C version from the authors' web-site[2]. We chose, however, to use a modified version of the PnET-II code (Radtke, 1999) because of familiarity within our lab and because it had been used for another study in Minnesota (Reich et al., 1999).

The process flow is diagrammed in Figure 2. The Perl 5.x scripting language was used extensively for data preparation and proved to be an invaluable tool in integrating the various modeling and GIS software. With the capability to directly execute both models, Perl also provided an efficient means for large batch runs. Code for the primary Perl scripts used in this research are included in Appendix B.

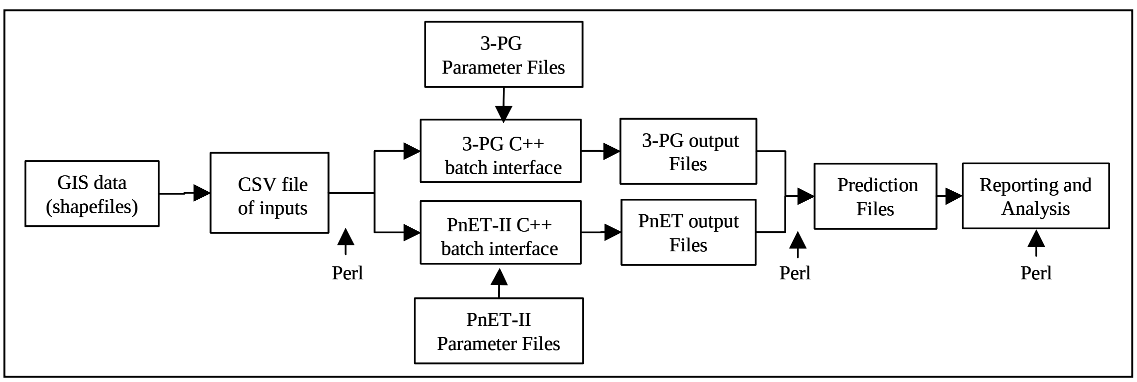


Figure 2: Process flow of model implementation. Sequences for which Perl was used are identified.

---

[1] 3-PG code available at: http://www.ffp.csiro.au/fap/3pg/
[2] PnET code available at: http://www.pnet.sr.unh.edu/

The model was run for each forest grid cell in the study area. This resulted in 25,885 runs of each model for the 1 km scale and 299 runs each for the 10 km scale. Additional batch runs were subsequently completed for sensitivity analysis (described below in section 2.5.5).

Model calibration is often the initial step for EPM studies. In this study, calibration runs were used in order to evaluate the estimated parameter values. However, the models were not calibrated against "real world" forest observations (although they were validated against other regional studies). It was hypothesized that these well-established models would replicate forest production in the region.

2.4.3 Sensitivity Analysis

Latin Hypercube Sampling (LHS) was selected as the sensitivity analysis technique for this project. LHS is an extension of stratified sampling for which all portions of the full range of each input variable are explicitly represented. LHS has been found to be a more efficient sampling method than random Monte Carlo techniques (McKay et al., 1979). The LHS algorithm (Figure 3) was implemented as described by Walters (1994) and written in Perl 5.x within the batch run script (code included in AppendixB). After all sensitivity runs were completed, the relative effects of the input were analyzed with correlation analysis and simple linear regression.

1) Select number of input configurations (i.e., number of runs), N
2) Identify the M variables of interest and the range of each variable (M_{min}, M_{max})
3) Calculate the interval width of each variable:

$$M_{width} = (M_{max} - M_{min}) / N$$

4) Randomly select combinations of intervals (M_{int}) for all variables of interest, and ensure that there are no duplicate combinations.
5) Assign values (M_{value}) to use in the model run for each variable. This requires identifying the value at the lower end of the selected interval, and then randomly selecting a value within the interval width:

$$M_{value} = M_{min} + (M_{int} * M_{width}) + (\text{rand}(0,1) * M_{width})$$

where rand(0,1) is a random number between 0 and 1
6) Run the model with each set of M_{value} combinations

Figure 3: Pseudo-code for the Latin Hypercube Sampling Algorithm.

2.4.4 FIA Validation Analysis

Additional calculations were required in order to compare modeled estimates of NPP against the FIA database. For this study, an area-weighted, regression-based approach (Brown and Schroeder 1999) was used to convert volume growth estimates from the FIA database into NPP production estimates. This approach uses a relationship between Growing Stock Volume (GSV, or the volume of commercial tree species) and Aboveground NPP based on empirical studies. Biomass Expansion Factors are calculated for each county, general forest type, and stand-size class combination, which are in turn applied to the net annual growth and mortality of GSV to convert them to estimates of biomass. NPP estimates are calculated as the sum of net annual biomass production and annual mortality. Once NPP estimates are calculated for each forest type/stand size combination, the estimates are aggregated to a county level using area-weighted averaging. Brown and Schroeder (1999) found significant differences for pine, spruce-fir, and hardwoods. Thus, for this analysis, the maple-basswood and aspen-birch groups are combined.

2.4.5 Spatial Analysis

Two methods were used to evaluate spatial patterns of forest production estimates. First, the Moran's I test for spatial autocorrelation was calculated. Second, variogram models were fit to the 1 km resolution estimates. These techniques provide methods for evaluating spatial patterns as well as comparing model estimates. Spatial autocorrelation is the intuitive notion that objects near each other are more likely to be similar than objects far apart. Moran's I quantifies spatial autocorrelation by using a weighted correlation coefficient to test for departures from spatial randomness (Cliff and Ord, 1981). The value for Moran's I falls between -1 and 1, where 0 equals spatial randomness and negative or positive numbers indicates the level of negative or positive autocorrelation, respectively. For this project, Moran's I was calculated using the S-Plus for ArcView 3.x extension (S-Plus, 1998).

The second spatial analysis tool used is a variogram model. A semi-variogram (commonly shortened to variogram) measures spatial autocorrelation by estimating the variance of observations as a function of the distance between the observations (Isaaks and Srivastava, 1989). A variogram is calculated as:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{(i,j)|h_{ij}=h} (v_i - v_j)^2 \qquad [2]$$

where $\gamma(h)$ = the semivariance for lag-distance *h*, *N* is the number of observations for lag distance *h*, and v_i and v_j are each pair of observed values *h* distance apart. The plotted variogram (figure 4) has the increasing lag distances on the X axis, and the variance on the Y axis. A variogram is useful for two reasons. First, it provides a visual tool for spatial data exploration of a continuous variable, similar to the use of histograms for univariate data. For example, a variogram supplies an estimate of overall population variance (the sill), the spatial extent of autocorrelation (the range) and an estimate of measurement error or bias (nugget effect). The second primary use of a variogram model is to define distance-based weights for a spatial interpolation method called kriging. This study used variograms exclusively as a spatial data exploration tool. All variogram analysis was done in the Surfer 8.0 software (Surfer, 2003).

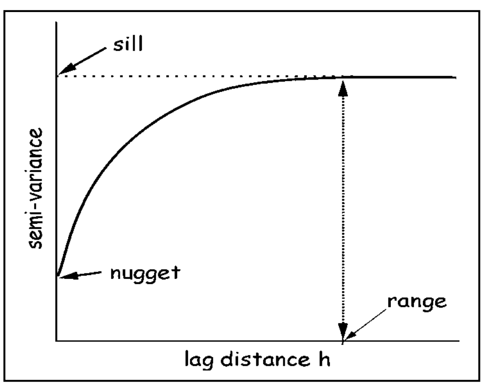


Figure 4: Sample variogram model (from Bolstad, 2002).

2.5 Results & Discussion

Four sets of model runs are compared, consisting of a 1x1 km and a 10x10 km set for each of the two models. Although both models provide a variety of outputs, only aboveground biomass growth estimates are evaluated. Aboveground NPP (ANPP) is the sum of foliage and stem production for the year (ANPP is shortened to NPP for this analysis). This limited analysis was chosen for two reasons. First, forest managers are usually more interested in above ground growth (particularly stem growth) than below ground growth. Second, the 3-PG model stores only cumulative component biomass pools, not annual production. Thus, modification of the 3-PG code would be required to extract component production estimates.

The mean NPP estimates for the four modeling sets ranges from 783.2 to 820.9 g C m^{-2} yr^{-1}, with standard deviations between 218.7 and 255.0 g C m-2 yr^{-1} (See Table 1 for all summary statistics). Thus, in the most general sense, all model predictions fall in a comparable range. More detailed analysis of spatial resolution effects within each model, inter-model comparison, spatial analysis and sensitivity analysis are presented below.

2.5.1 Summary Statistics

Summary statistics of the four model runs are included in Table 1. Figure 5 contains histograms of the 1x1 km and 10x10 km NPP estimates. In aggregate, the data are multi-modal at both resolutions with no discernable pattern. This is explained by the forest type, for which the models are parameterized. Separating by cover type results in unimodal distributions for all cover types except for the 10x10 km Aspen-Birch group, which is bimodal. Overall, none of the cover type predictions are normally distributed (chi-square test for normality, $p < 0.001$). As a result, non-parametric tests are used for comparison of estimates.

Table 1: Univariate data summary for PnET-II and 3-PG model runs.

	NPP Estimates (g C m^{-2} yr^{-2})			
Statistic	**PnET 1k**	**PnET 10k**	**3-PG 1k**	**3-PG 10k**
Overall				
Count	25885	299	25885	299
Minimum	478.5	425.1	444.1	504.3
Maximum	1250.0	1097.9	1106.0	1173.5
Mean	842.8	783.2	788.8	820.9
Median	1018.7	832.1	951.6	823.0
Standard Deviation	255.0	218.7	246.6	233.8
Coef. Of Skewness	-0.34	-0.30	-0.33	-0.17
Pine				
Count	1720	15	1720	15
Minimum	707.6	706.8	673.7	615.4
Maximum	836.0	784.0	819.2	803.3
Mean	783.2	757.9	729.1	713.5
Median	788.0	760.9	734.0	705.7
Standard Deviation	17.9	19.3	18.9	77.0
Coef. Of Skewness	-1.15	-1.16	-0.23	-0.001
Spruce-Fir				
Count	9599	97	9599	97
Minimum	478.5	425.1	444.1	504.3
Maximum	561.2	527.4	529.8	549.3
Mean	524.9	494.3	482.7	527.5
Median	525.0	492.6	480.0	527.5
Standard Deviation	9.4	19.2	20.8	9.26
Coef. Of Skewness	0.02	-0.52	-0.79	-0.05
Aspen-Birch				
Count	13658	181	13658	181
Minimum	964.4	723.5	924.1	712.5
Maximum	1250.0	1097.9	1106.0	1173.5
Mean	1064.0	936.6	1003.4	984.5
Median	1062.9	935.0	1001.2	1050.2
Standard Deviation	33.6	99.3	37.7	123.7
Coef. Of Skewness	0.60	-0.18	-0.32	-0.91
Maple-Basswood				
Count	908	4	908	4
Minimum	767.8	855.9	728.3	806.43
Maximum	1046.6	918.5	968.7	1010.7
Mean	989.3	889.0	911.1	898.1
Median	1008.8	889.2	915.6	862.8
Standard Deviation	55.1	25.3	44.1	85.8
Coef. Of Skewness	-2.12	-0.09	-2.43	0.70

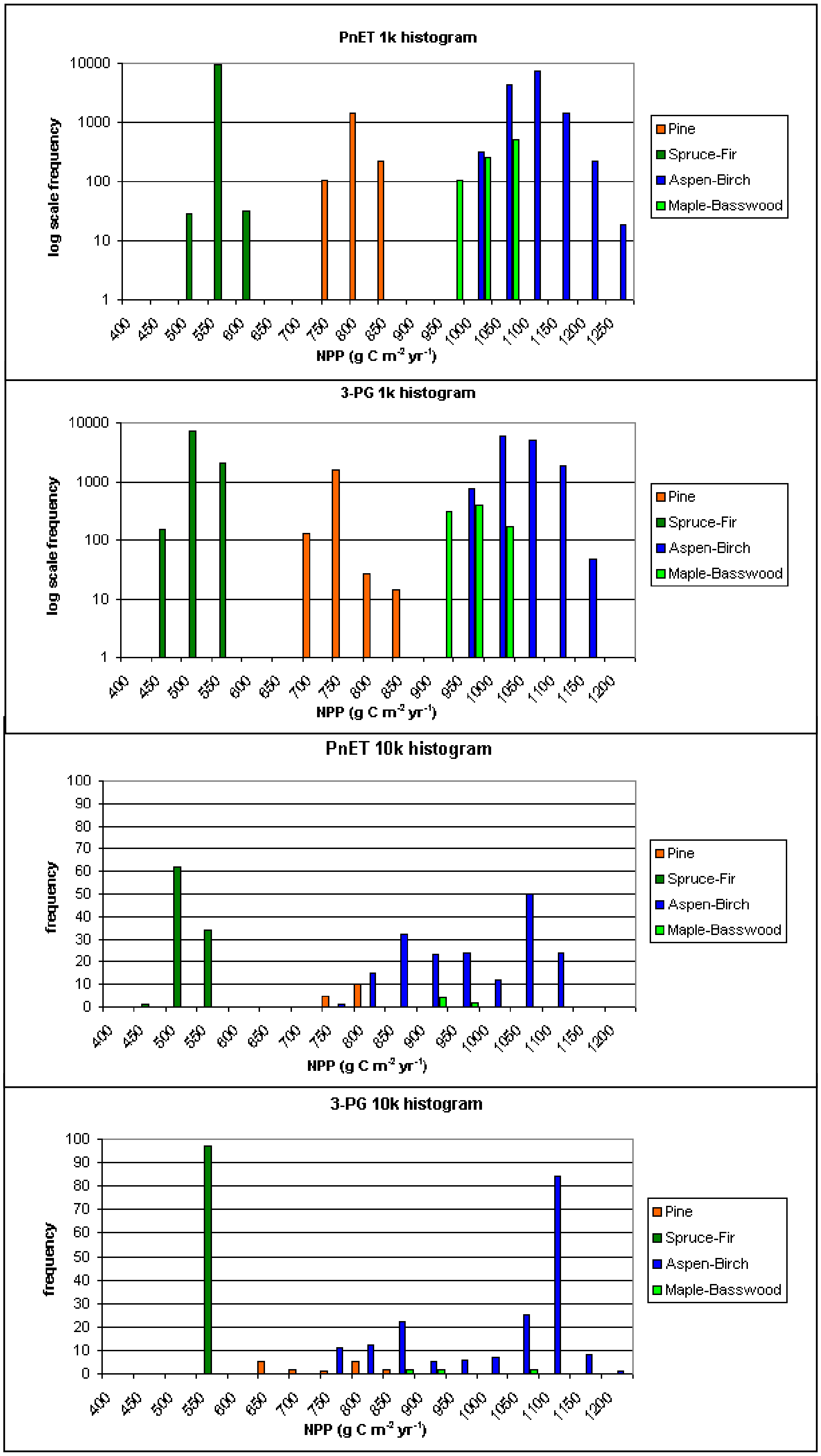


Figure 5: Histograms for prediction estimates by cover type. The top two graphs are for the 25885 1x1 km cells, with log scale frequency. The bottom two graphs are for the 299 10x10 km cells.

2.5.2 Inter-model Comparison

Figure 6 provides a scatterplot of the model predictions for the two spatial scales. In general, the model predictions are in relative agreement. However, there are some notable differences. At the finer spatial resolution (1x1 km), PnET-II predictions are significantly higher than the 3-PG estimates for all cover types (Wilcoxon Rank-Sum Test, $p<0.001$, Table 2). For the 10x10 km resolution, however, there are no systematic patterns across all cover types. The scatterplots also illustrates the relative variance in estimates between the two models. For the equilibrium PnET-II model, the variation within each cover type is relatively low, while the non-equilibrium 3-PG estimates vary over a much wider range. Furthermore, if the forest age class distributions were accounted for in the 3-PG model, the variation would be even larger.

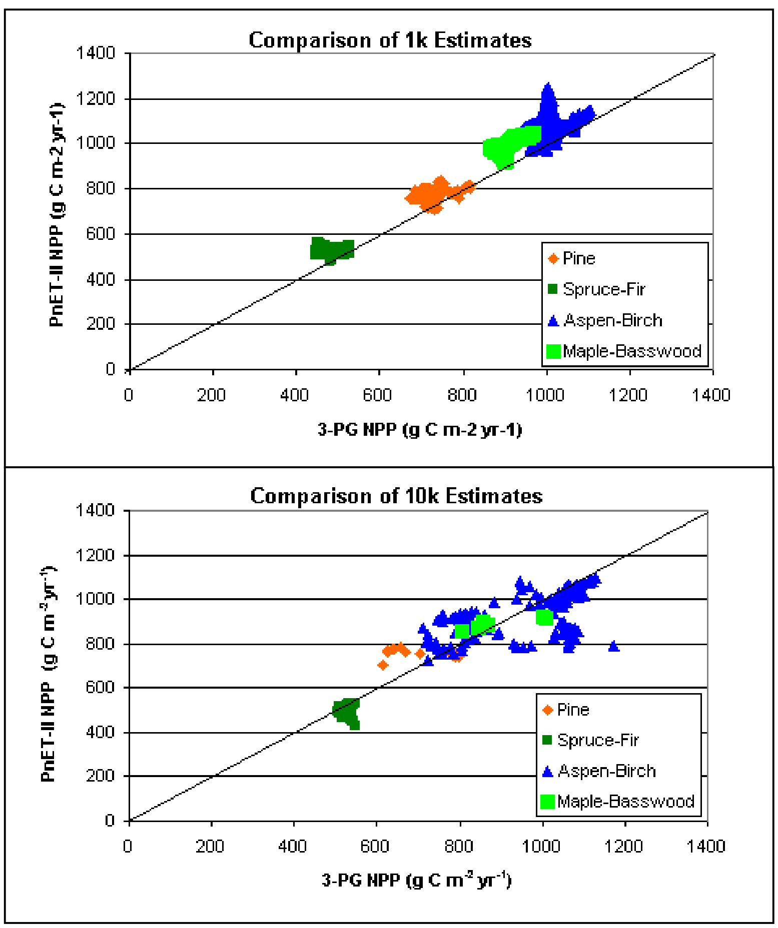


Figure 6: Scatterplot of model estimates for the 3-PG and PnET-II models. The top figure is for the 25885 1x1 km cells, and the bottom for the 299 10x10 km cells. The diagonal lines indicate a 1:1 ratio.

Table 2: Comparison of model predictions. Z-scores and p-values are based on Wilcoxon Rank-Sum tests between the estimates at the same scale for the two models.

Cover Type	Overall		Pine		Spruce-Fir		Aspen-Birch		Maple-Basswood	
1x1 km Resolution										
No. Cells	25885		1720		9599		13658		908	
Model	PnET-II	3-PG	PnET-II	3-PG	PnET-II	3-PG	PnET-II	3-PG	PnET-II	3-PG
Mean (Std. Dev)	842.8 (255.0)	788.8 (246.6)	783.2 (17.9)	729.1 (18.9)	524.9 (9.4)	482.7 (20.8)	1064.0 (33.6)	1003.4 (37.7)	989.3 (55.1)	911.1 (44.1)
Z-Score (p-value)	190.9 (< 0.001)		47.6 (< 0.001)		109.0 (< 0.001)		108.2 (< 0.001)		30.9 (< 0.001)	
10x10 km Resolution										
No. Cells	299		15		97		181		4	
Model	PnET-II	3-PG	PnET-II	3-PG	PnET-II	3-PG	PnET-II	3-PG	PnET-II	3-PG
Mean (Std. Dev)	783.2 (218.7)	820.9 (233.8)	757.9 (19.3)	713.5 (77.0)	494.3 (19.2)	527.5 (9.26)	936.6 (99.3)	984.5 (123.7)	889.0 (25.3)	898.1 (85.8)
Z-Score (p-value)	21.1 (< 0.001)		120 (< 0.001) *		12.0 (p < 0.001)		5.5 (< 0.01)		21 (< 0.001) *	

* For the two 10x10 km cover types with less than 20 observations, the Exact Wilcoxon Rank-Sum Test is used. For those two, reported scores are for the "w-scores", not the "z-scores".

2.5.3 Scale Analysis

On a per-meter basis, increasing spatial resolution has a significant effect on model predictions. That is, mean predictions were significantly different between the 1x1 km and 10x10 km predictions for both models and all cover types (Wilcoxon Rank-Sum Test, $p < 0.001$, Table 3). In all cases, the 10x10 km resolution NPP estimates were lower than the 1x1 km estimates. Aggregating data and removing heterogeneity likely mitigated or removed the extreme scenarios that would lead to higher model predictions. This may also be explained by the aggregation methods. In this project, continuous data were aggregated using a straight averaging function while categorical data were aggregated with a block-majority rule. Thus, heterogeneity was lost through averaging. Alternative approaches, such as weighted averaging, might help retain heterogeneity.

Table 3 shows the effects of aggregation on cover type distributions. The block-majority aggregation rule resulted in a relative increase in the most prevalent cover type (Aspen-Birch) and a relative decrease in all other cover types. In this study, Aspen-Birch was the most productive cover type as well as the most prevalent. In regions where the most prevalent cover type is not the most productive, the effects of aggregation may be even more pronounced.

Table 3: Comparison of scale effects. Z-scores and p-values are based on Wilcoxon Rank-Sum tests between the 1x1 km and 10x10 km estimates for the same model.

	Cover Type									
	Overall		**Pine**		**Spruce-Fir**		**Aspen-Birch**		**Maple-Basswood**	
Spatial Resolution	1x1 km	10x10 km	1x1 km	10x10 km	1x1 km	10x10 km	1x1 km	10x10 km	1x1 km	10x10 km
Number of Cells	25885	299	1720	15	9599	97	13658	181	908	4
% Area	100	100	6.6	5	37.1	32.4	52.8	60.5	3.5	1.3
PnET-II Mean (Std. Dev)	842.8 (255.0)	783.2 (218.7)	783.2 (17.9)	757.9 (19.3)	524.9 (9.4)	494.3 (19.2)	1064.0 (33.6)	936.6 (99.3)	989.3 (55.1)	889.0 (25.3)
Z-Score (p-value)	29.78 (< 0.001)		6.68 (< 0.001)		16.97 (< 0.001)		18.09 (< 0.001)		4.22 (< 0.001)	
3-PG Mean (Std. Dev)	788.8 (246.6)	820.9 (233.8)	729.1 (18.9)	713.5 (77.0)	482.7 (20.8)	527.5 (9.26)	1003.4 (37.7)	984.5 (123.7)	911.1 (44.1)	898.1 (85.8)
Z-Score (p-value)	29.34 (< 0.001)		5.46 (< 0.001)		5.93 (< 0.001)		4.72 (< 0.001)		2.14 (0.032)	

An interesting effect of aggregation in this analysis is that there were no systematic changes in the patterns of variations within and between cover types. For example, aggregating the Spruce-Fir cover type led to an increase in variation for the PnET-II predictions, but a decrease in variation for the 3-PG predictions. The reverse pattern is true for the Maple-Basswood cover type. This suggests that aggregation may have non-uniform and non-linear effects between the mean and variance. Thus, it appears unsafe to assume that aggregating data will have a uniform and linear influence on model predictions.

2.5.4 Spatial Analysis

Figure 7 provides maps of the NPP predictions for the two models and two spatial scales. From visual observation, it appears that the data are spatially autocorrelated. Some general trends are discernable. Areas of lower productivity are found in the northwest section of the study area. This makes sense considering the extensive wetland systems in the area where forest growth is limited. Much of the Spruce-Fir is located in this area. Conversely, the area of highest productivity is located in the south central portion of the study area that contains the more productive upland forests such as Aspen-Birch and Maple-Basswood.

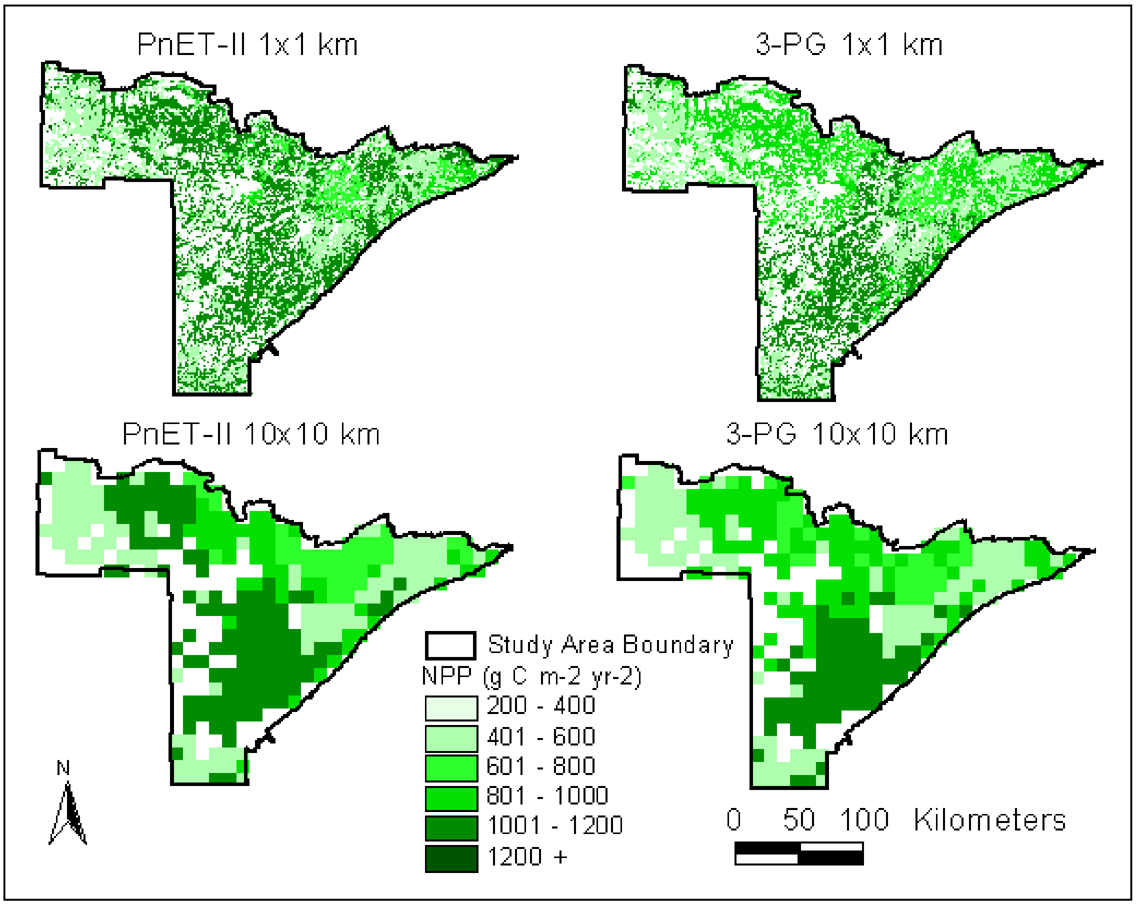


Figure 7: NPP predictions for the two models and two spatial resolutions.

Two measures of spatial autocorrelation were calculated: Moran's I for both models and resolutions, and variograms for the 1x1 km scenarios. Variograms were not calculated for the 10x10 km scenarios because of the relatively low ratio between the cell width (10 km) and the overall spatial extent (360 km). The Moran's I autocorrelation index values indicate a positive spatial autocorrelation for all four modeling scenarios, with the two 1x1 km scenarios having a similar and higher autocorrelation than the 10x10 scenarios (Table 4). The positive autocorrelation matches a visual interpretation of the NPP maps in Figure 7. However, the decrease in autocorrelation with the aggregated data is surprising, given the clumping on the maps (although the clumping may be a function of the categorical grouping method). This decreased autocorrelation could possibly be due to the large perimeter and relatively high number of no-value cells of the coarse data.

Table 4: Moran's I spatial autocorrelation index values for the two models and resolutions.

Model	Spatial Resolution	Moran's I Statistic
PnET-II	1x1 km	0.58
	10x10 km	0.32
3-PG	1x1 km	0.56
	10x10 km	0.32

The variogram analysis also indicates similar spatial patterns between the two models (Table 5, Figure 8). Both show spatial autocorrelation over a range of up to 65 km. However, both models show the peculiar pattern of having the variogram model level off and then rise again (Figure 8). Thus, the selected exponential model with nugget effect is not a very good fit. No better model forms were found. The second rise in variance may be a result of the clumping of similar values in the study area due to the large influence of cover type. It may also be due to the relatively small range of NPP estimates (450-1200 g C m-^{2} yr^{-1}) over the region.

Table 5: Variogram parameters for the 1x1 km model runs

	3-PG NPP	PnET-II NPP
Model	Exponential with nugget effect	Exponential with Nugget Effect
Sill	56000	63000
Nugget Effect	38000	45000
Maximum Range	65,000 m	59,000 m
Minimum Range	20,500 m	21,000 m
Anisotropy Ratio	2.5	2.5
Direction of Max Continuity	60^{o} north of east	60^{o} north of east
Direction of Min Continuity	30^{o} south of east	30^{o} south of east

Both variograms have an anisotropic (i.e., directional) pattern, which was identified by calculating variograms in different directions and comparing the ranges. The direction of maximum continuity is in a north-northwest direction. This matches the primary geologic patterns in the region, including the hilly range that extends along the Lake Superior shoreline. All together, both models appear to have similar spatial patterns of NPP predictions.

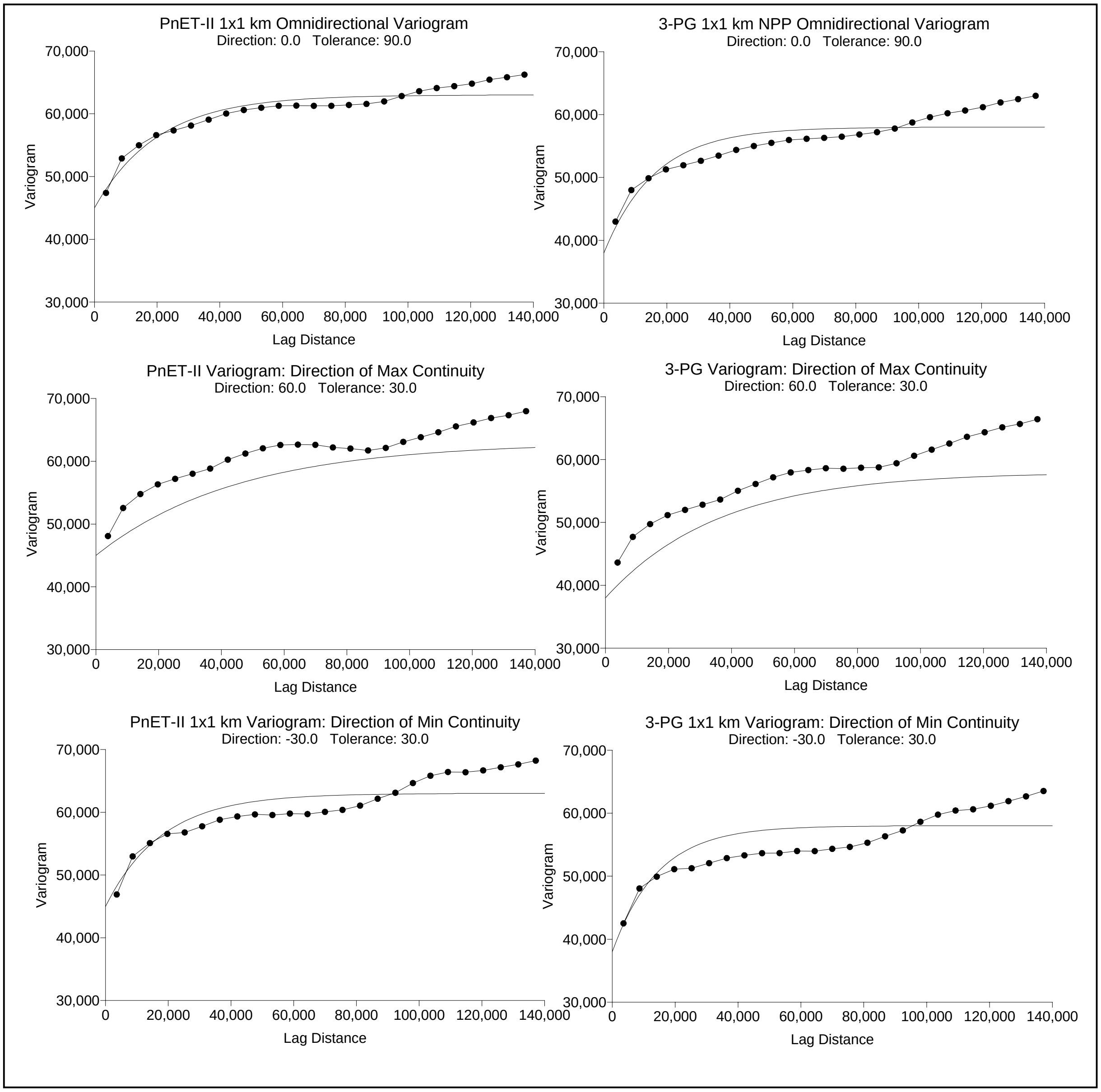


Figure 8: Variogram models for the 1x1 km runs. PnET-II variograms are on the left and 3-PG variograms are on the right. The top row contains omnidirectional variograms. The middle and bottom rows contain variograms for the direction of maximum and minimum continuity, respectively. The dotted line represents the mean paired observations variation for each lag distance. The solid line is the omnidirectional variogram model fit described in table 5.

2.5.5 Sensitivity analysis

Sensitivity analysis was conducted for 100 randomly selected grid cells across all cover types. For both models, three input variables were chosen for sensitivity analysis: SWHC, temperature and precipitation. Sensitivity analyses are commonly performed on both parameters and input variables. For this project, however, the objective was only to compare common input variables. Values were chosen uniformly from the range of observed values in the region for SWHC (35 – 750 mm). For temperature, all monthly values were varied up to 5 degrees C from observed values. For precipitation, monthly values were varied up to 5 cm. These values were selected to represent a large range of variation over the course of the year. Monthly values were not varied individually because of the very large increase in processing that would be required.

The first round of analysis consisted of 100 repetitions, with each variable altered individually. Simple correlation analysis was conducted on the change in the variable against the change in baseline NPP from the main model runs (Table 6).

Table 6: Correlation coefficients of changes in individual input variable vs. change in predicted NPP.

	Correlation Coefficient (R)	
Variable	**PnET-II**	**3-PG**
Temperature	0.84	0.71
Precipitation	0.13	0.20
SWHC	0.21	0.07

The second analysis consisted of varying all three variables simultaneously using Latin Hypercube Sampling (see section 2.4.3). Regression analysis was used to evaluate the relative influence of each variable on the overall change in NPP from the baseline values. Surprisingly, temperature was the only significant variable ($p < 0.0001$) when all three variables were included in the regression fit. A simple linear regression fit of change in temperature on change in predicted NPP resulted in an $R^2 = 0.62$ for the 3-PG run and $R^2 = 0.70$ for the PnET-II runs.

Overall, this sensitivity analysis illustrates the importance of understanding the influence of various inputs on model outputs. In this study, temperature was by far the most important input variable evaluated. As such, special care should be taken in developing the temperature data set. For this case study, a simple inverse-distance weighted (IDW) interpolation technique was applied to the long-term normal monthly temperature values for climate stations in the region. IDW interpolations typically produce a “bullseye” effect (e.g., see sample maps in Appendix A), with a sharper change in interpolated values nearer to sample locations. Looking back, other interpolation techniques might have been preferable, such as a regional regression approach.

2.5.6 Validation

Model validation was not performed for any specific location within the study area. Because our objectives focus on regional forest production estimates, we chose to only compare modeled estimates against those from other studies in the region and against aggregated estimates from the FIA database. Table 7 compares NPP predictions for previous studies within or near our study area. In all cases but one, both the PnET and 3-PG estimates fell below the upper range of estimates from the other studies. However, in all cases, the minimum estimates from the other studies were well below the minimum estimates from this study. Thus, the predictions in this study may have an upward bias on NPP estimates.

Table 7: Comparison of published NPP estimates against estimates from this study.

		NPP Estimate Range (g C m^{-2} yr^{-1})		
Cover Type	**Source**	**External**	**PnET**	**3-PG**
Research studies				
Pine	Fassnacht & Gower, 1996	390 – 850	707 – 836	615 – 819
Maple-Basswood	Fassnacht & Gower, 1996	290 – 1150	768 – 1047	728 – 1011
Spruce-Fir	Hall et al., 1992	40 – 572	425 – 561	444 – 549
Aspen-Birch	Hall et al., 1992	190 – 1199	724 – 1250	713 – 1174
FIA-based studies				
Hardwood species	Brown & Schroeder, 1999	410 – 800	724 – 1250	713 – 1174
Softwood species	Brown & Schroeder, 1999	210 – 600	425 – 836	444 – 819
Pine	Recalculated from Brown & Schroeder 1999	151 – 569	707 – 836	615 – 819
Spruce-Fir	Recalculated from Brown & Schroeder 1999	185 – 438	425 – 561	444 – 549
Hardwood species	Recalculated from Brown & Schroeder 1999	259 – 1197	724 – 1250	713 – 1174

Another form of validation is to compare NPP estimates against estimated growth between the FIA inventory years. As described above in section 2.4.4, Brown & Schroeder (1999) used a regression-based approach to estimate NPP from the FIA database. Table 6 also contains these county-level aggregated mean estimates for the 5 counties in our study area, as well as the range of values found by reapplying the Brown and Schroeder calculations to the 1977 and 1990 Minnesota FIA survey results. As with the research studies, these FIA-based data suggest model estimates may have a large upward bias. This is most evident in the Pine and Spruce-Fir cover types where there is no overlap between estimate ranges for all cases but one. The regression coefficients used in Brown and Schroeder (1999) and Schroeder et al. (1997), and reapplied in this study, were generalized for the entire Eastern U.S. Thus, they may not be appropriate for the finer scale of analysis in this study.

Compared against all other studies combined, it is apparent that our study fails to capture scenarios of lower biomass production. Low NPP may be caused by several factors such as site conditions and forest age structure. In this study, a representative age range of 30-60 years was used. Estimates may be improved modeling the full range of age distributions in the study area rather than by relying on this representative age range. Similarly, site variables such as SWHC were derived from generic data sources. In all, improvements in data quality and detail may likely improve the overall forest production estimates in the region.

2.6 Conclusions

One of the most significant observations of this study was the large effect forest cover type had on production estimate variation. Particular care should be made in both mapping the cover type and in parameterizing the models for each cover type. This is especially true for the equilibrium PnET-II model. For the 3-PG model, we elected to only evaluate the general production estimates for common stands in the middle of their rotation age. By more accurately incorporating forest age structure, production estimates would likely result in a much wider distribution among and within cover types.

Another area of concern is in the use of GIS data from a variety of sources and with a large range of data quality. The scale analysis indicated that aggregation has a variety of effects that are not systematic. For this study area, aggregating to a coarser resolution resulted in lower mean estimates, uncertain effects on variability, and decreased spatial autocorrelation. However, these effects appeared to be tied to the cover type distribution, for which the most prevalent cover type (Aspen-Birch) was also the most productive. Other studies have shown conflicting effects. For example, Turner et al. (2003) found a similar pattern to ours for coniferous landscapes in the Pacific Northwest. They concluded that capturing heterogeneity by increasing spatial resolution resulted in increased estimates of Net Ecosystem Production (NEP) compared with coarser resolution modeling. Conversely, McNulty et al. (2000) concluded that a spatially-explicit version of PnET run for southern forests was insensitive to aggregation of input variables, but concluded that aggregations can bias estimates. For this research, it was acceptable to have such uncertainty because the focus was on the methods and not the predictions. For applied research and operational use, emphasis should be placed on evaluating data quality, and sensitivity analysis should be conducted on all input layers in addition to the parameter values. If possible, a multi-scale analysis is suggested in order to help identify potential sources of error or bias.

With these concerns in mind, there were also encouraging trends across all model runs. First, it was reassuring that the overall predictions from the two models were at least comparable, although this may be a function of parameterizing one of the models (3-PG) against the other (PnET-II). The fact that the models were within the range of estimates from previous studies gives support to the argument that the models are comparable. Similarly, scale analysis and spatial autocorrelation analysis were similar between the two models. Thus, if properly parameterized, multiple EPMs can be used in a similar fashion. It is common practice in many studies to compare predictions from more than one model. If feasible, a multiple-model approach is also suggested.

In terms of model implementation, both models were run within the Perl-based "wrapping" approach. This suggests that, even with varying underlying theory and software packages, it is possible and relatively straightforward to overcome such differences with current technologies. One downside to this approach, however, is increased computational costs. In this project, the biggest computation bottleneck resulted from having to format input ASCII files, feed them into the models, and then reformat output ASCII files. This separation of code and reliance on ASCII text files clearly slowed processing down. For example, a batch run for each of the 25885 cells in the 1x1 km analysis took over 12 hours on a 1.0 GHz Pentium III processor. Obviously, this time will be reduced with faster desktop computers. However, processing times could be drastically reduced with current computers if the code were more fully integrated. In this research, though, the flexibility and simplicity of using the Perl "wrapper" approach was well worth the computational cost.

3 A general framework for implementing regional-scale ecosystem models

3.1 Introduction

The use of computer-based forest simulation models has increased dramatically over the past two decades. Scientists and forest managers use these tools to address practical and hypothetical scenarios across all spatial scales, from predicting growth of individual trees to calculating worldwide carbon budgets under global warming conditions. A large amount of research has focused on regional-scale forest modeling (e.g., watershed or ecoregion scale). Studies at such coarse scales rely heavily on GIS and remote sensing technologies. The increase in technology, complexity, and scope of these modeling studies has necessitated the development of methods for implementing and using the models, and standards for administering the large amounts of data required to run the models.

A major criticism of coarse-scale modeling studies is that few people understand the needs, processes or results of the studies. The technical implementation is often to blame for this criticism. Model development and data administration practices are often selected on an ad hoc basis with little or no concern for repeatability or longevity. This chapter directly addresses the issue of defining a modeling framework. The objectives of this chapter are to:

1) develop a generalized conceptual framework for regional-scale Ecosystem Process Models (EPMs),

2) evaluate various technical modeling approaches for EPMs, and

3) identify future trends in the use of EPMs.

The overarching goal of this project is to evaluate the process of implementing regional-scale EPMs and identify future trends and opportunities. Chapter 1 provided background information. Chapter 2 consisted of a sample implementation of two popular EPMs for the Arrowhead region of Northeastern Minnesota. This chapter uses the knowledge gained in Chapters 1 and 2 to critically evaluate the modeling process.

3.2 Conceptual Framework

Forest growth models have often been implemented for independent, single-use studies. Although data and the core model are regularly reused, the modeling process flow, technological setup and analyses are often recreated for each separate modeling study. There are, however, several identifiable model implementation components that are common across studies. For example, studies typically contain data preparation, parameter calibration, sensitivity analysis and reporting phases.

For this framework, we identify four general components of model implementation. First, there is the model itself, which is considered here to be a “black box” model for which the inputs and outputs are the primary concern, not the internal model algorithms. Second, model implementations require data sources and data stores for both inputs and outputs. Third, technological processing is required in order to manage data and convert between formats. Finally, there are ecological model considerations, which include the scientifically based methods and decisions generally included with model studies. Figure 1 contains a conceptual flowchart for a traditional “model-centered” modeling framework based on these four components.

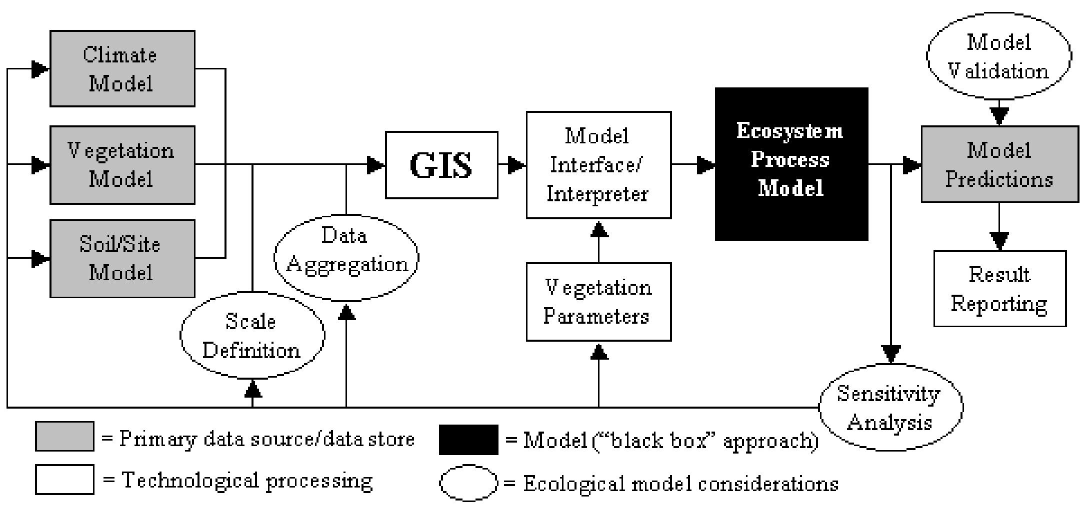


Figure 1: Conceptual framework of an EPM implementation interfaced with a GIS.

While EPMs may be encapsulated in a single program, this framework is based on the argument that a modeling study is much broader than the model itself. Key decisions need to be made long before the model is ever run, and the output of the first model will likely never be accepted as the definitive answer. The remainder of this chapter explores important technological, data administration, and model assessment topics relating to the broader scope of EPM studies as defined in the conceptual model presented in figure 1.

3.3 Technologies

Computer-based technologies provide both endless opportunities and large challenges for estimating forest production. A limiting factor frequently identified as a weakness of current EPM implementations is the information system organization. While one lab or user may have developed a modeling environment that works well for them, that environment is often not easily transferred to other groups or scaled up for use in a larger organizational context. As a starting point in improving the information system organization of EPMs, this section explorers the various technological components of the EPM conceptual framework.

3.3.1 GIS

A fundamental decision in spatially explicit, process-based modeling is the relationship between the model and the GIS. The conceptual flowchart presented in Figure 1 above illustrates a traditional "model-centered" framework. This approach requires the user to incorporate the various components, understand and properly address the ecological model considerations, and properly validate and report the results, and is arguably the most common approach in scientific modeling studies. Mitasova and Mitas (2002) identify this approach as an "import/export" (or "loose coupling") interface between the model and GIS (Figure 2). Two alternatives include an “embedded coupling,” or full integration of the model within the GIS, and a “tight coupling,” which provide of a common interface between the GIS and model.

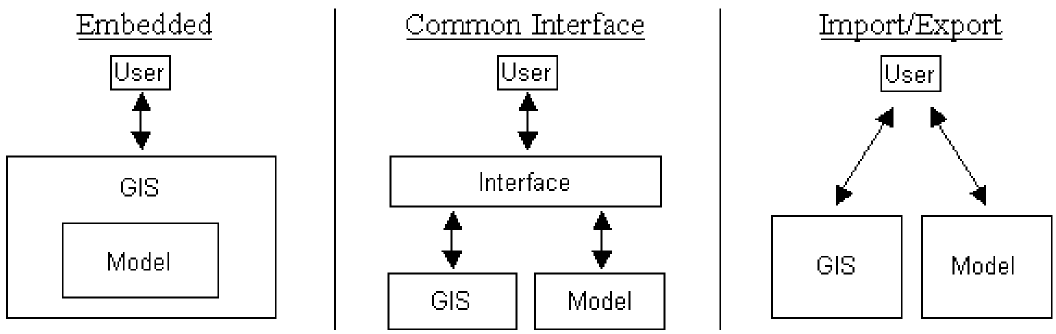


Figure 2. Visualizations of three categories of integration between models and GIS.

Overall, there is a trade-off between end-user friendliness of the more fully integrated alternatives and development flexibility of the less integrated alternatives. With recent advances in GIS software capabilities, such as the Microsoft COM-compliant structure in both ESRI’s ArcGIS software suite and IDRISI’s Kilimanjaro package, it is becoming easier to integrate models within a GIS. Conversely, popular scripting languages such as Perl and Python can be used as relatively simple and flexible “wrapper” programs in which to link the various model components in a less rigid arrangement. Thus, the technological tools are available to meet a variety of GIS-related modeling approaches.

3.3.2 Remote sensing

Remote sensing technologies are almost always a direct or indirect component in EPM studies when applied across broad areas. Plummer (2000) identified four strategies for using remote sensing and remotely sensed data in EPM studies. First, remotely sensed data can be used to estimate variables required in EPMs (i.e., model inputs), such as LAI or forest type. Second, remote sensing can be used to test or validate predictions of EPMs. For example, EPM predictions can be compared against simple, reflection-based models of NPP. Third, remote sensing can be used to update or calibrate EPM predictions. For example, a regression between sparse field-based and extensive remotely sensed estimates of LAI can be used to calibrate final LAI estimates used in a model study. Finally, EPMs can be used to aid the interpretation of remotely sensed data. For example, this inverted approach can be used to evaluate empirical NDVI based estimates of LAI using more detailed and mechanistic understanding incorporated into EPMs.

The use of remote sensing in EPM studies will likely continue to increase. Turner et al. (2004) argue that remote sensing and process-based models are complementary technologies that, when combined, can provide an effective means to evaluate C cycles across large areas. With the large variety of new satellite and airplane sensor systems currently in use or in development, this area of research appears to be growing in the foreseeable future.

3.3.3 Climate models

Climate variables such as temperature and precipitation are invariably key inputs in EPMs. Typically, such variables are collected from climate stations in the region and spatially interpolated for areas where stations are not located. A variety of interpolation methods may be used. Ollinger et al. (1995) use a statistical approach dependent on latitude and longitude. Bolstad et al. (1998) concluded that regional regression interpolation provided better predictions for temperature than local lapse rates and kriging-based interpolations in mountainous areas. As part of the sample implementation used in this research project, a standard Inverse Distance-Weighted interpolation method was used (see Chapter 2, section 2.5.5). Spatial interpolation is a broad and important spatial problem with no clear best methods for all circumstances.

In terms of implementation, climate models are most often used to estimate the requisite variables prior to the actual model runs. For sensitivity analyses, two approaches are possible. First, the climate model can be directly incorporated into the sensitivity analysis sequence, effectively nesting the climate model within the ecological model. Second, a separate routine can be used to vary the values. For example, a script can be written to alter temperature by 5% above and below the value generated by the climate model. Climate variables often are the key drivers for light-use efficiency style EPMs. As such, climate data should be used carefully in modeling studies regardless of the interpolation technique or usage within the modeling framework.

3.4 Data administration

Process-based forest growth models require a large variety of data inputs. Before running a modeling study a significant amount of pre-processing is required to convert the data into a form suitable for the study. Running the model with inaccurate or incomplete data can have undesirable consequences on the outcome. Additional data challenges arise in a multi-user situation. Controlling access and updates becomes a requirement. Identifying user needs and abilities and providing appropriate interfaces is important. Improved data administration is one of the fundamental requirements for extending EPMs from the research lab to the production environment. The topics of data modeling, collection and aggregation are evaluated in this section.

3.4.1 Data modeling and entity relationships

Understanding the relationships and organization of data is a key first step in effective data administration. Having a stable and efficient data schema may help minimize potential biases and inaccuracies in an EPM study. A generalized logical data schema for regional-scale forest growth modeling studies is described in this section.

For forest systems, Denkers (1992) suggests modeling data based on subject entities and associations rather than application-driven structures, an approach commonly applied in forest growth modeling. A logical data scheme should be developed by rigorously defining and verifying entities, determining the relationships between entities, and formally creating a physical data scheme. As a result, the final database model will be more stable and flexible.

For fine-scale forest growth modeling (e.g., individual trees in a forest stand), Robinson (1998) proposed the use of an object-oriented data model, with Trees, Climate, and Soil being the primary objects. This approach does not apply as easily to coarse-scale forest growth modeling. First, the definitions of objects become less clear as scale is generalized. Second, coarse-scale modeling studies traditionally rely on GIS data structures, which in most cases involve relational data structures. Thus, a different logical model is often necessary.

Landsberg and Gower (1997) define four general categories of input requirements for EPMs: site conditions, vegetation/stand conditions, weather conditions, and vegetation parameters. These categories provide a starting point for defining entities in coarse-scale ecosystem modeling. Site conditions consist of data describing the area where the stand exists. Examples of applicable attributes are slope, aspect, soil texture, and latitude. Stand condition data define the condition of the forest stand itself, such as age, standing biomass, tree density, and LAI. Weather conditions data contain all pertinent climatic attributes, such as precipitation, incoming solar radiation, and temperature. Lastly, vegetation type parameters are defined as values specific to each vegetation cover type that describe how that cover type responds to environmental stimuli. For example, one common parameter is maximum photosynthetic rate, which is a value that differs among tree species and forest types.

These four general input data classes do not, however, cover all data entities necessary for forest growth modeling studies. Additional data are required for the processing unit for which the model is run (i.e., the spatial extent or area used). This is typically an individual forest stand or, more commonly, a raster grid cell within a GIS. An example of a potential Entity Relationship Diagram based on these classes is in Figure 3. In general, mapping data relationships in such a manner allows for improved understanding of input requirements, a structure for systematically updating and maintaining the database, and a framework for organizing the overall modeling study.

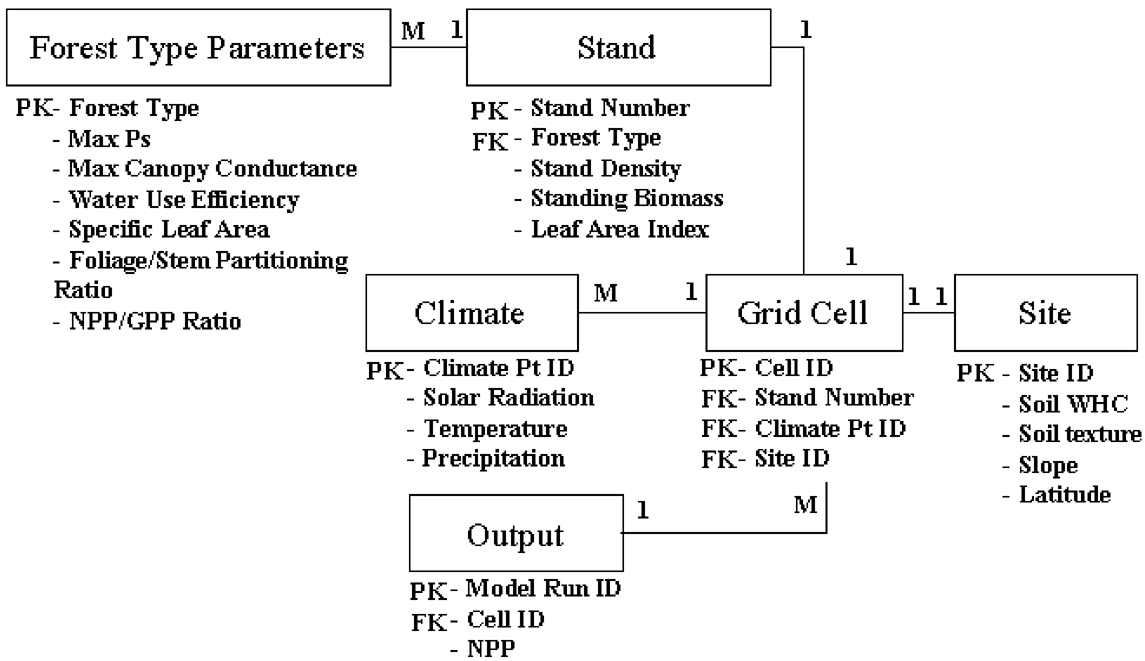


Figure 3: Sample Entity Relationship diagram for organizing data used with an EPM.

3.4.2 Parameter collection and estimation

Most process-based models are highly dependent on the parameter value set used in the study. Parameter values are obtained in a variety of ways. In some larger modeling studies, field-based experimental or observational studies are commissioned to determine particular parameter values. For parameters that are common across the scientific discipline, estimated values can be obtained from searches of the scientific literature or based on expert opinion. In other cases, parameter values are estimated using a calibration process applied to a sample data set such as a pilot study. A variety of statistical and software tools have been developed to aid the process. For example, Radtke (2002) used a Bayesian synthesis approach for identifying distributions for parameter values from which the models can be run. Commercial software is also available. For example, PEST (S.S. Papadopulos and Associates, Inc. 2004) is a model-independent parameter estimation and uncertainty analysis package used across many scientific and engineering disciplines. Overall, a wide assortment of resources are available to aid in the parameter estimation process.

3.4.3 Data scaling and aggregation

For any forest modeling study, a unit of interest must be defined for which the model is run. In regional-scale modeling, this unit is most often a raster grid cell. Much effort has been spent in Geographic Information Science and related fields in evaluating the effects of using different cell sizes on process outputs. For example, Kang et al. (2004) modeled C and water cycles across 8 different spatial resolutions and found that model outputs varied nonlinearly with spatial aggregation of inputs, with biases of up to 50% for one key variable (solar radiation). The sample implementation in this research project also found non-linear and non-uniform patterns resulting from aggregation (see Chapter 2 section 2.5.3).

Clearly, defining the spatial resolution(s) used in analysis is a very important decision. In general, there is a trade-off between computation time and capturing heterogeneity in the system being modeled. There is also a trade-off between the generalization errors inherent in aggregating data and potential classification errors in using fine-grained data. In order to specifically address this problem, many studies now run models across a variety of resolutions to quantify errors associated with aggregation. Regardless, the resolution of analysis should be careful considered in implementing modeling studies and interpretation of results.

3.5 Model assessment

Model evaluation consists of the methods and tools used to assess whether or not a model is properly performing the tasks for which it was designed. With the huge amount of data that EPMs typically provide as output, it is important to critically and quantitatively evaluate those data. Radtke (2001) argues that there is a discrepancy between the sophistication of methods used in process-based model development and the tools for the evaluation of those models, especially when compared against the broad evaluation methodologies for traditional statistical models. Inferences derived from modeling studies can be interpreted more confidently with the use of more powerful and robust evaluation methodologies. Two general categories of model assessment are discussed here: sensitivity analysis and calibration. Methods for evaluating and estimating parameter values were mentioned in section 3.4.2 above.

3.5.1 Sensitivity Analysis

Sensitivity analyses can be classified into of two broad categories. In the first, varying scenarios are predefined, with the model runs for each scenario. This is the often the case with "what-if" analyses, for example, what if atmospheric CO_2 concentrations doubled? In the second approach, a large number of repetitions are conducted with random or systematic variations applied to the desired input or parameter variables. In this case, the objective is either to evaluate the relative importance of various model inputs in determining model outputs, or else to statistically explore the relationships within and among model inputs, parameters, or processes being modeled. The simplest, and arguably most common, approach is Monte Carlo sensitivity analysis, in which one or more input variables are randomly altered with each repetition. More advanced methods include Latin Hypercube Sampling (McKay et al. 1979; see Chapter 2 section 2.4.3 for a further description), Fourier Amplitude Sensitivity tests (Radtke 1999), and fractional factorial analysis (Henderson-Sellers and Henderson-Sellers 1996; White et al. 2000). Each of these methods propounds statistical or computational improvements over basic Monte Carlo analysis.

3.5.2 Calibration, Validation, and Uncertainty Assessment

In any study that predicts forest growth, regardless of the modeling approach, a key requirement is determining how well the predictions mimic the true forest growth in the field. With traditional statistical approach, a model is fit (i.e., calibrated) against a sample data set. Uncertainty estimates are provided via standard error statistics or confidence intervals. The model can then be validated against a separate sample data set. Due to the increased complexity, these tasks can be much more challenging for EPMs, but are just as important.

In practice, EPMs are often validated in a variety of ways. First, predictions can be compared against separate data sets not used in model development or calibration. Second, predictions can be compared to other studies in the same area of interest. Third, predictions can be compared against other modeling approaches, such as remote sensing studies.

If field-based studies or similar data are available, the models may be calibrated against those data. With broad-area studies, however, thorough calibration against field-based studies may be logistically or financially impractical. In this case, the distinctions between calibration and validation are blurred. Approaches traditionally considered being validation techniques are used to calibrate the model during initial pilot model runs. Similarly, uncertainty assessment of model predictions is restricted by small or nonexistent sample data sets, resulting in the need for surrogate approaches such as using sensitivity analyses to identify plausible confidence limits. See Radtke (2001) for a fully developed discussion of these topics.

3.6 Conclusion and Future Trends

This project used one particular modeling implementation in northeastern Minnesota as the basis for identifying a framework for implementing regional-scale forest production models. Using the PnET-II and 3-PG models with publicly available data sets, forest production estimates were found to be comparable to estimates from other studies in the region in terms of production estimates and spatial patterns. The four general components of the modeling framework include the model, technological processing steps, data sources and stores, and ecological modeling considerations. Overall this study supports the general consensus of experts as found in a review of the literature that there are large potential benefits for expanding EPMs into operational forest management environments.

Turner et al. (2004) identify future trends in regional-scale use of EPMs. Particularly, advances in remote sensing technologies have expanded the set of pertinent data that can be used with EPMs. In addition to vegetation type and LAI, new approaches are being developed to estimate standing biomass, forest age structure, canopy height, and canopy chemistry properties, such as foliar leaf Nitrogen. Conversely, EPMs are being developed or modified to take advantage of these new information sources. Lucas and Curran (1999) argue that remote sensing can provide the basis for multi-scale simulation of forest ecosystem processes. In addition to providing a source for mapping forest cover type, remote sensing can be used to estimate forest composition, biophysical, physiological, and biochemical properties.

Landsberg (2003) lists several current challenges and research trends in the use of process-based models in operational environments. Some of the key challenges in current models are a lack of scientific understanding in the processes of C allocation within trees, nutrient availability and cycling in the soil, and uptake of nutrients by the trees. Landsberg also cites advances in the linkage between remote sensing and EPMs as a current research trend.

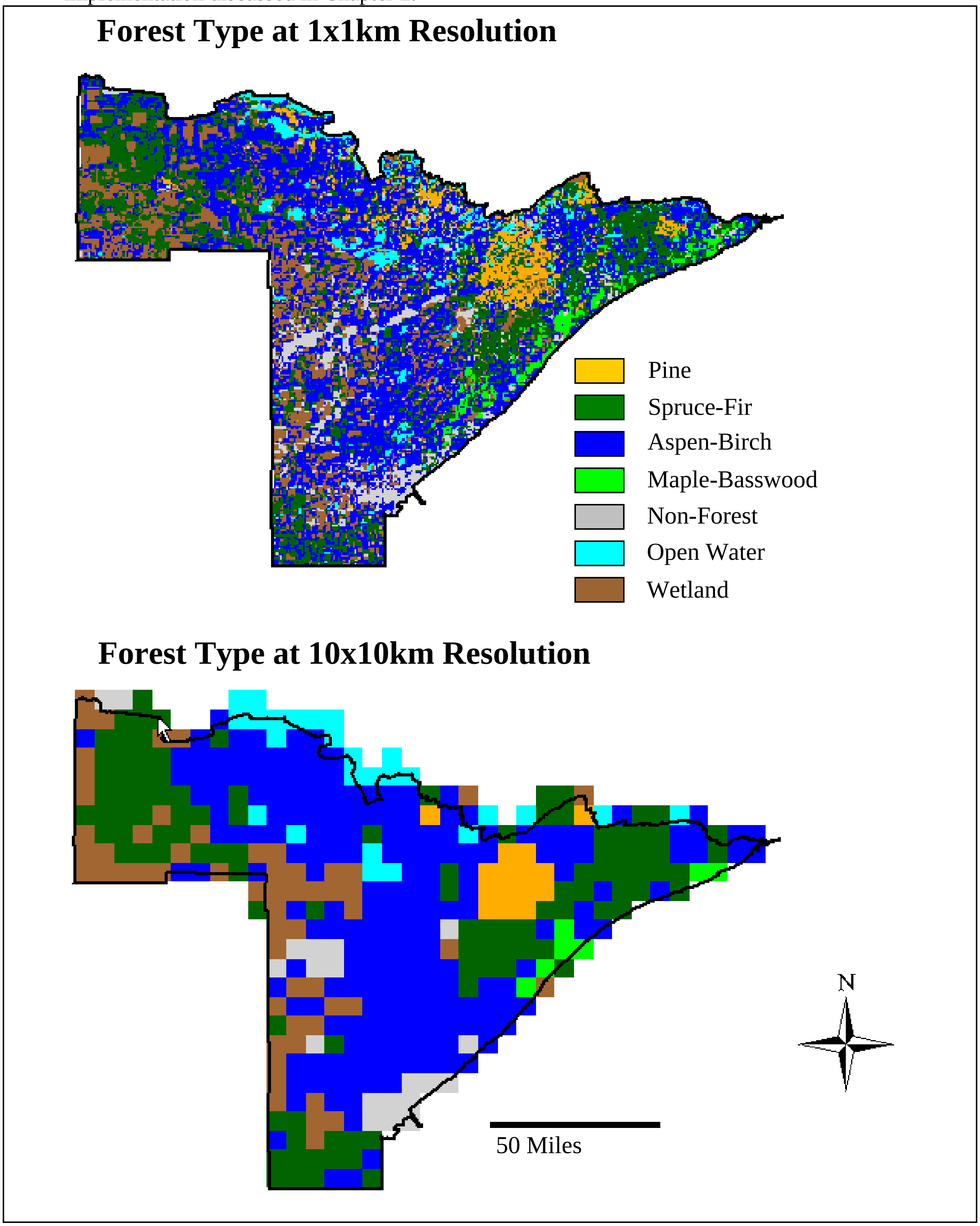
A third set of current trends consists of expanded and novel modeling strategies. For example, Liu et al. (2002) developed a component object model for “reusing” existing ecosystem models as modules in a larger model. More generally, Landsberg (2003) notes that models also benefit from expanded use of the World-Wide Web. Communication is increased, updates can be distributed more quickly, and data and parameter sets can be shared.

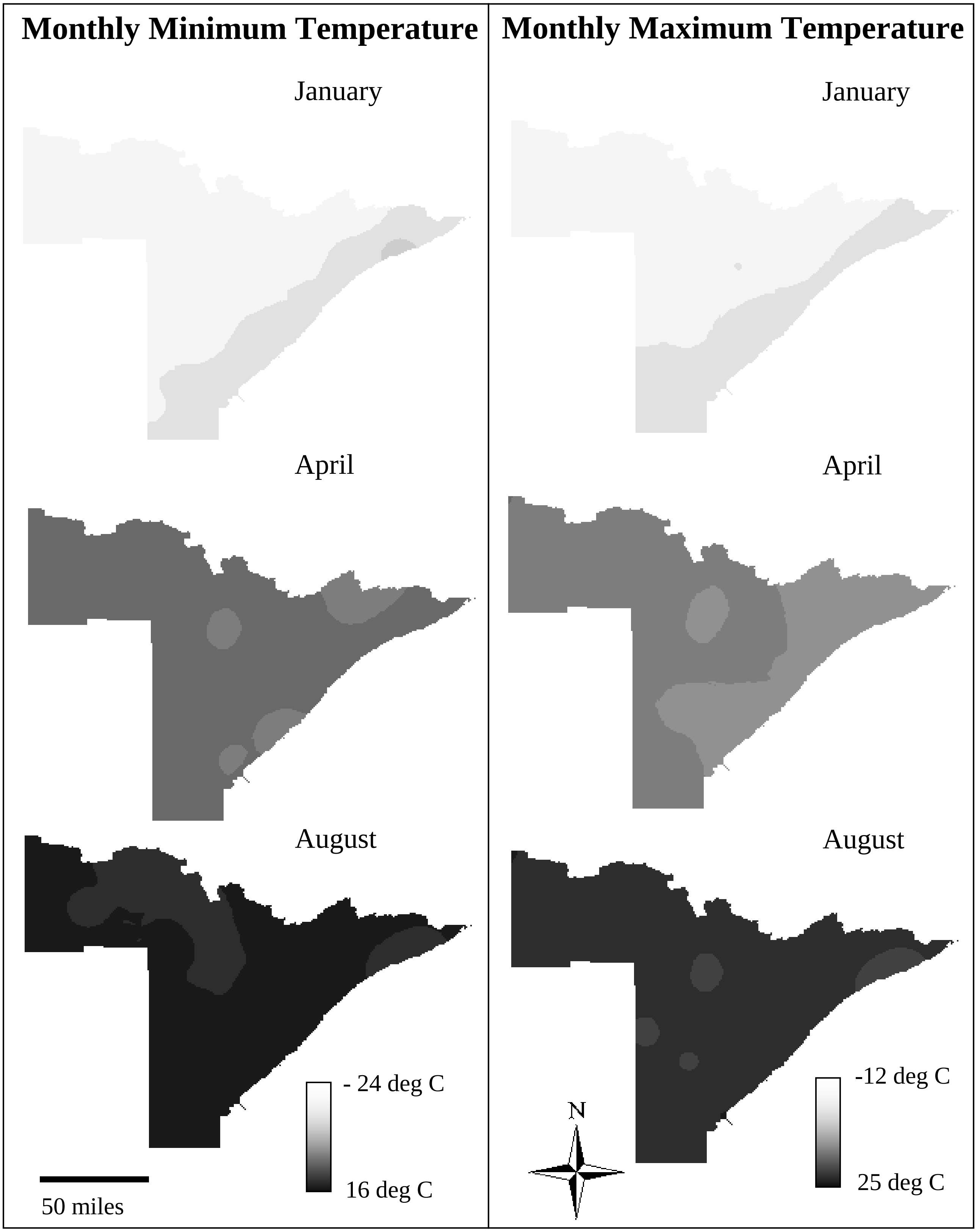
Besides these technological challenges and trends, it is apparent that specialized training is required to properly implement process-based models in a production environment. This includes general understanding of the ecological processes being modeled, specific knowledge of the computer-technologies being used, and scientific comprehension of the statistical, analytical, and scaling issues surrounding process-based models. Over the past several years university-level courses, public short courses, and publications have been developed to explore the alternatives and consequences of these modeling considerations. With improved model implementations and proper training, which is becoming increasingly available, a myriad of new management modeling analyses are possible.

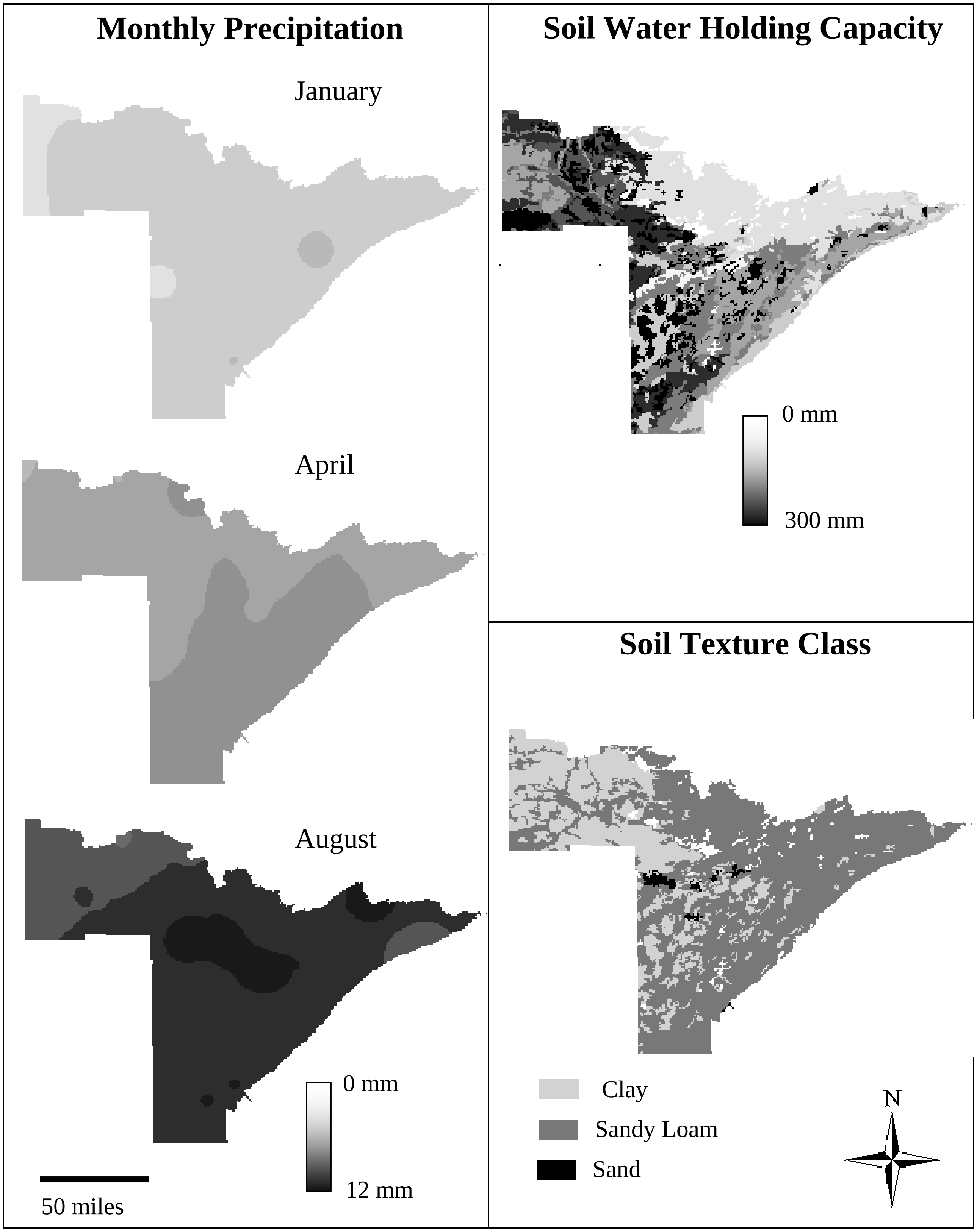
To date, there is only one documented example of the 3-PG model being used in an operational environment (Almeida et al., 2002), and no known operational uses of PnET-II have been identified. Indeed, process-based models have yet to find a widespread niche in traditional forest management. In his review article, Landsberg (2003) identifies the primary challenge and opportunity for the practical use of EPMs is the need for communication between model developers and forest managers. Addressing the information system and GIS components within a modeling framework may provide a common ground for expanding this communication.

Appendix A: Sample Maps

This appendix contains sample maps of the primary input data sets used in the sample implementation discussed in Chapter 2.




Monthly Minimum Temperature
Monthly Maximum Temperature
January
January
April
April
August
August
- 24 deg C
16 deg C
50 miles
N
-12 deg C
25 deg C


Monthly Precipitation
January
April
August
0 mm
12 mm
50 miles
Soil Water Holding Capacity
0 mm
300 mm
Soil Texture Class
Clay
Sandy Loam
Sand
N

Appendix B: Technological implementation of models

This appendix provides a description and sample code of the approach used in this study for implementing regional scale models using GIS data. The approach consists of using the Perl scripting language as a "wrapper" around the model being used. This approach provides a link between the GIS data set and the model, and allows for batch processing of the executable files of the two models used. Figure 1 illustrates the general flow of how the models were used and linked.

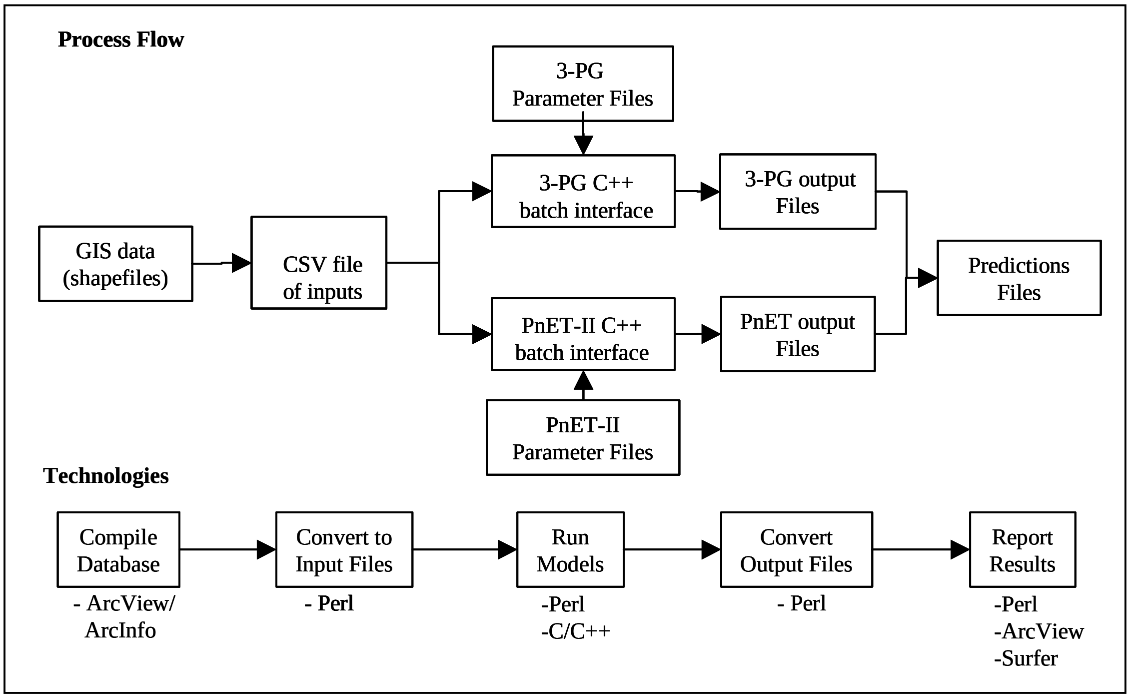


Figure 1: Process and Technological flow of model implementation.

All input requirements were organized as attributes in a point feature Shapefile, with each point representing the center of the grid cell for which the models are run. Once completed, the attribute table was exported to a Comma Separated Value (CSV) file. A Perl script was written to read each line of the CSV file, convert the data into the format used by the model, and run the model for that cell. Separate Perl scripts were used to sort, summarize, and analyze the output files.

The primary Perl script used for running the PnET-II model is included here as reference. The code for 3-PG is different, but follows the same organization, so is not included here. The Perl scripts also contain the code for the optional Latin Hypercube Sampling sensitivity analysis. (comments begin with a pound sign, #).

```
#!\c:\perl\bin -w

#This script controls a batch run of the PnET-II model.
#As inputs, it requires three types of files:
#  1) Input variable csv table, often exported from an ArcView
#     shapefile .dbf file.
```

```
#  2) A two-column csv file, with each record containing
#     (a) a variable name as listed in the input variable file
#     (b) the corresponding variable name used in PnET
#     (this allows for different naming schemes).
#  3) PnET-II Parameter files for each vegetation type
#
#To run, use <name>.pl <input var file>,
#where <name> is the name of this perl script and
#<input var file> is the file name of file 1) described above

use Benchmark;

$time1 = new Benchmark;

$sitFileLoc = "d:/sims";
$outDir = "output-pnet";
$pnetVarList = "pnetvarlist.txt";

#hash of cover types used in this batch run, with the first item
#being the numeric code, and the second being the name. The names
#have to be the same as used in the parameter files.
%goodCovType = (1 => "Pine", 2 => "S-F", 3 => "A-B", 4 => "M-B");

#dates to run the model.
$startYear = 1977;
$endYear = 2027;

#month mid points, in julian days, used by PnET
@midMonth = (15,46,76,107,137,168,198,229,259,290,321,351);

#if running sensitivity analysis, fill these values in.
#Currently, only 4 variables are set up for sensitivity
#analysis, folN, SWHC, temp, and precip. folN and SWHC
#are based on the range of observed limits.  Temp and Precip
#are based on a percentage variation of long term known
#averages. All sensisitivity analysis code will have to
#be changed for varying additional parameters.
#
#The sensitivity analysis uses latin hypercube sampling,
#where all variables are varied together from the given
#distributions.
$numN = 100;     #number of runs for each processing unit
                 #set numN to 0 if not doing sensitivity
                 #analysis
$folNmin = 1.7;  #min value for foliarN
$folNmax = 2.7;  #max value for foliarN
$SWHCmin = 1.5;  #min value for Soil Water Holding Capacity
$SWHCmax = 29;   #max value for SWHC
$tempmin = -5; #lower range of % variation of temperature
$tempmax = 5;  #upper range of % variation of temperature
$precmin = -5; #lower range of % variation of precipitation
$precmax = 5;  #upper range of % variation of precipitation

#Set up sensitivity analysis if applicable:
if ($numN > 0) {

 #find the interval width to use in sampling.
 $wfolN = ($folNmax - $folNmin) / $numN;
 $wSWHC = ($SWHCmax - $SWHCmin) / $numN;
 $wTemp = ($tempmax - $tempmin) / $numN;
 $wPrec = ($precmax - $precmin) / $numN;

 #find random numbers of each value within its given
 #range for each run and put them in arrays. Also,
 #make sure there are no duplicate combinations of
 #randoms numbers for any given run.
 push(@r1,int(rand $numN));
 push(@r2,int(rand $numN));
 push(@r3,int(rand $numN));
 push(@r4,int(rand $numN));
 for ($b = 2; $b <= $numN; $b++) {
   do {
     $rand1 = int(rand $numN);
     $rand2 = int(rand $numN);
     $rand3 = int(rand $numN);
     $rand4 = int(rand $numN);
     $duplicate = 0;
     for($c = 0; $c <= $#r1; $c++) {
       if ($rand1==$r1[$c] && $rand2==$r2[$c] && $rand3==$r3[$c] && $rand4==$r4[$c]) {
         $duplicate = 1;
         $c = $#r1;
```

```
      }
    }
  } until ($duplicate == 0);
  push(@r1, $rand1);
  push(@r2, $rand2);
  push(@r3, $rand3);
  push(@r4, $rand4);
 }

 #Assign actual values to use model runs for each variable and
 #place them in an array of length NumN.
 for ($d = 0; $d < $numN; $d++) {
  $r1val = $folNmin + ($r1[$d] * $wfolN) + (int(rand) * $wfolN);
  $r2val = $SWHCmin + ($r2[$d] * $wSWHC) + (int(rand) * $wSWHC);
  $r3val = $tempmin + ($r3[$d] * $wTemp) + (int(rand) * $wTemp);
  $r4val = $precmin + ($r4[$d] * $wPrec) + (int(rand) * $wPrec);
  push(@folNvector,$r1val);
  push(@SWHCvector,$r2val);
  push(@tempvector,$r3val);
  push(@precvector,$r4val);
 }
} # end sensitivity analysis setup


#############################################
### Part 1: Map input fields to PnET vars ###
#############################################

#
### Read header line of input file to get array of field names
#
$line = <>; #this reads from the input file 1) described above
chomp $line;
@head = split(",",$line);

#
### load PnET variable list, file 2) as described above
#
open(pnetVarList, "< $pnetVarList") or die "no open $pnetVarList";
while ($varRec = <pnetVarList>) {
 chomp $varRec;
 @inputRec = split(",", $varRec);
 push(@pnetVars, $inputRec[0]);
}

#
### For each PnET variable, find the column location
### in the input file where that variable is located
#
for($i=0;$i<=$#pnetVars;$i++) {
  for($j=0;$j<=$#head;$j++) {
    if ($pnetVars[$i] eq $head[$j]) {
      $pnetLoc[$i] = $j;
      last; #this stops the $j loop after each match
    }
  }
}

#############################################################
### Part 2:  Create input files for each record; run PnET ###
#############################################################

while (<>) {
#
### read record in
#

chomp;
@record = split(",",$_);

##########
### create Pnet climate (clm) and site (sit) file.
##########

#
### first, put required data values into an array using
### the mapped location array
#

for ($k=0;$k<=$#pnetVars;$k++) {
#  print "k = $k, pnetloc = $pnetLoc[$k], rec = $record[$pnetLoc[$k]] \n";
```

```
  $fields[$k] = $record[$pnetLoc[$k]];
}

#if doing sensivity analysis, create and open summary file
if($numN > 0) {
  $finalFile = $fields[1] . "-" . $fields[0] . "-sens.csv";
  open(finalCellList, "> $outDir/$finalFile") or die "can't open finalFile $finalFile";
  print finalCellList "folN,SWHC,temp-perc,rain-perc,NEP,NPPFolYr,NPPWoodYr,NPPRootYr
\n";
}

#start loop if doing sensitivity analysis, otherwise will
#just run once.

for ($a=0;$a<$numN;$a++) {

  if ($numN > 0){
    #assign for sensitivity analysis
    $folN = $folNvector[$a];
    $SWHC = $SWHCvector[$a];
    $tempm = $tempvector[$a];
    $rainm = $precvector[$a];
#print "$a, $folN,$SWHC,$tempm,$rainm\n";
  } else {
    #assign default values
    $folN = 2.0;
    $SWHC = 12.0;
    $tempm = 0;
    $rainm = 0;
  }

  ### Create PnET Site File ###
  #$fields[1] is the cover type code
  #$fields[0] is the unique ID
  #$fields[2] is the latitude
  $sitFile = $fields[1] . "-" . $fields[0] . ".sit";
  open(writeFile, "> $sitFile") or die "no open sit $sitFile";
  print writeFile "File Directory **********" . "\n";
  print writeFile $sitFileLoc . "\n";
  print writeFile "SiteVariables ***********" . "\n";
  print writeFile "LAT     WHC     Climate file \n";
  print writeFile $fields[2] . "   " . $SWHC . "   " . $fields[1] . "-" . $fields[0] .
"\n";
  print writeFile "Initial Conditions ******" . "\n";
  print writeFile "BudC  WoodC  PlantC  NRatio  PlantN  FolMass  WoodMass  RootMass \n";
  print writeFile "0     300     900    " . $folN . "    1      0     47000     6 \n";
  print writeFile "SnowPack Dwater Water   HumusM   HumusN  NH4   DeadWood \n";
  print writeFile "  0        1     0       13500    390     .01    11300  \n";
  print writeFile "Scenario - for CN *******" . "\n";
  print writeFile "Run Model  From/To \n";
  print writeFile " $startYear    $endYear \n";
  print writeFile "Run Climate File From/To \n";
  print writeFile " $startYear   $endYear \n";
  print writeFile "DelTMax DelTmin DelPrec  DelPar  DelWUE   Ramp?   Start   End \n";
  print writeFile "    0      0       1       1       1       0       0       0 \n";
  print writeFile " WetNO3  WetNH4  DryNO3  DryNH4   Ramp?   Start     End    Bkgd \n";
  print writeFile " .28999     .13  .20399     .05       1    1940    2000     .25 \n";
  print writeFile "FertNO3  FertNH4 YrStart  YrEnd  MonStart MonEnd \n";
  print writeFile "     0       0       0       0       0       0  \n";
  print writeFile "FolRegen \n";
  print writeFile "    100 \n";
  close (writeFile);

  ### create PnET Climate File ###
  $clmFile = $fields[1] . "-" . $fields[0] . ".clm";
  open(writeFile, "> $clmFile") or die "no open clm $clmFile";
  $numYears = $endYear - $startYear;
  $numMonths = $numYears * 12;
  print writeFile "$numMonths \n";
  print writeFile "Year  DOY  Tmax    Tmin     PAR   Prec    NH4      NO3     O3 CO2 V1
V2 V3 V4\n";

  #loop through tmax, tmin, PAR, and precip for each month of each year for climate.
  #using long term averages, so will be repetitive. $tempm is for sensitivity analysis.
  $clmSt = 5; #this is the starting point of the climate files -- may need to be changed;
  for ($y = 1; $y <= $numYears; $y++) {
    $x = $clmSt - 1;
    for ($z = 1; $z <= 12; $z++) {
      $tTMin = sprintf( "%.3f",$fields[$x] + $tempm);
      $tTMax = sprintf("%.3f", $fields[$x+12] + $tempm);
      $tRain = sprintf("%.3f", $fields[$x+36] + $rainm);
```

```
      print writeFile "$y    " . $midMonth[$z-1] . "  " . $tTMin . "  " . $tTMax . "  "
. $fields[$x+24] . "  " . $tRain . " 0.00   0.00   0.00   0.00 \n";
      $x++;
    }
  }
  close (writeFile);

  #
  ### Run PnET model for cell ###
  #
  $vegFile = $goodCovType{$fields[1]} . ".veg";
  $outputFile = $fields[1] . "-" . $fields[0] . "-" . $numYears . ".txt";
  system "d:/sims/pnet_main $sitFile $vegFile $clmFile d:/sims/ > ./$outDir/$outputFile";
  unlink("$sitFile");
  unlink("$clmFile");
  #
  ### add to single list for sensitivity analysis
  #
  if ($numN > 0) {
    open(addFile, "< output-pnet/$outputFile") or die "can't open output file
$outputFile";
    do { $addLine = <addFile> } until $. == $numYears || eof;
    chomp $addLine;
    @ad = split(/\s+/,$addLine);
    print finalCellList $folN . "," . $SWHC . "," . $tempm . "," . $rainm . "," . $ad[2]
. "," . $ad[3] . "," . $ad[4] . "," . $ad[5] . "\n";
    close (addFile);
    unlink("$outDir/$outputFile");
  }

  $totalCells++;
  if ($totalCells/100 == int($totalCells/100)) {
    print "$totalCells ";
  }

} #end a loop for sensitivity analysis

#close sensitivity analysis file
if ($numN > 0) {
  close(finalCellList);
}

} # end main while <> loop


#####################
### Print Results ###
#####################

$time2 = new Benchmark;
$td = timediff($time2, $time1);
print "time: " . timestr($td) . "\n";
print "$totalCells processing cells generated for PnET-II. \n";
```

Appendix C: Conference Proceedings Paper

This appendix contains a reproduced copy of the paper submitted and presented to the 4th Southern Forestry and Natural Resources GIS Conference. The full citation is:

Kirk, Ryan W., and Thomas E. Burk. 2004. Regional-Scale forest production modeling using process-based models and GIS. In: SoFor GIS 2004: Proceedings of the 4th Southern Forest and Natural Resources Geographic Information Systems Conference. Athens, GA, December 2004.

REGIONAL-SCALE FOREST PRODUCTION MODELING USING PROCESS-BASED MODELS AND GIS

Ryan W. Kirk and Thomas E. Burk
Department of Forest Resources, University of Minnesota,
St. Paul, MN 55108

ABSTRACT

While research scientists have used process-based models of forest growth for several decades, forest managers have only recently begun to adopt them in production environments. This lag is accredited to the nature of process-based models, which are often difficult to parameterize, challenging to validate, and built around limited technical implementations. This study addresses these limitations by incorporating standard information system and GIS concepts into the modeling framework. As a sample implementation, the PnET-II and 3-PG models were run within a GIS for the Arrowhead region of northeastern Minnesota and compared against growth estimates from other studies in the region. Based on the experiences of this modeling study and a review of the literature, a framework for implementing process-based models within a GIS is presented. Primary components of the framework include ecological modeling considerations, data sources and stores, and technological processing requirements. Several GIS-based modeling strategies are evaluated. In addition, current technological and programming trends, research priorities and implementation challenges are discussed.

KEYWORDS. Process-based models, 3-PG, PnET, Minnesota, GIS Model

INTRODUCTION

Process-based models can be defined as formalized statements of hypotheses regarding a complex system and its responses to stimuli (Landsberg 1986). To scientists, such models are tools that provide a structure for organizing current knowledge of a particular system, a framework with which to test hypotheses about that system, and a means to evaluate responses to stimuli within the system (Landsberg and Gower 1997). Due to the complexity of calculations, process-based models are most often presented as stand-alone computer programs or nested within a spreadsheet or other software application. Thus, endless combinations of user interfaces, output, presentation, and analysis options are possible.

In recent years, forest managers have expressed interest in the application of process-based models in forest management decision making (Mäkelä et al. 2000, Korzukhin et al. 1996, Johnsen et al. 2001). In a summary paper, Battaglia and Sands (1998) identify five potential uses of process-based forest productivity models as management tools: (1) prediction of growth and yield, (2) selection of new plantation sites, (3) identification of site limitations on productivity, (4) assessment of risks associated with locations or management options, and (5) use of models as surrogates for field experiments. Mohren and Burkhart (1994) argue that process-based models provide greater potential for predicting forest growth under varying environmental conditions than empirical growth and yield models. From this perspective, the focus of modeling shifts away from scientific inquiry to strategic and operational considerations.

In the summary report of an International Union of Forestry Research Organization working group on applications of process-based models (Mäkelä et al. 2000), the group affirmed the potential of process-based models as management tools across all spatial scales. A primary recommendation of the group was for improved practical implementations within operational management systems. In essence, a key factor to further use of process-based models as decision-making management tools is in information system design and analysis, not just continued scientific development. Similarly, Battaglia and Sands (1998) argue that current process-based models are overly complex for practical use and are in a state of constant development; this is in disagreement with the desire for robustness and consistency in forest planning methodology (Sievänen and Burk 1993). Johnsen et al. (2001) contend that process-based models are quite valuable in simulating extremely complex forest systems, but will only be adopted when the complexities of research models are overcome. Korzukhin et al. (1996) conclude that with an increasing focus on ecosystem-based forest management, process-based models become a valuable tool for addressing a large variety of management decisions. Thus, with a diverse and clearly defined interest in process-based modeling for natural resource decision-making, additional efforts need to be made to join the needs of forest managers with the powerful models being developed by researchers.

This study aimed to provide support for bridging the gap between scientific and operational implementations of process-based forest production models. The primary objectives were to:

1. Implement a regional-scale modeling study using GIS and remote sensing technologies for northeastern Minnesota.
2. Compare regional predictions of Net Primary Production (NPP) using two popular process-based models (PnET-II and 3-PG) at two spatial resolutions.
3. Based on a literature search and the experiences and results of objectives 1 and 2, develop a generalized framework for the technical implementation of regional-scale process-based forest production models.

METHODS

Study Area

The study area is a 5 county area in northeastern Minnesota commonly called the "Arrowhead" region (Figure 1). The region is geologically identified as the Superior Uplands portion of the Canadian Shield, and is characterized by exposed bedrock, extensive mineral deposits, varied topography, thin soils, and a large abundance of lakes. The area is bordered on the east by Lake Superior and on the north by Canada. The climate is characterized by cold winters, cool summers and an average of 66 cm of precipitation per year. The region falls in a transition zone between temperate and boreal forests, containing predominantly aspen, birch, and white, red, and jack pine in the uplands, and spruce, fir and tamarack in the lowlands.

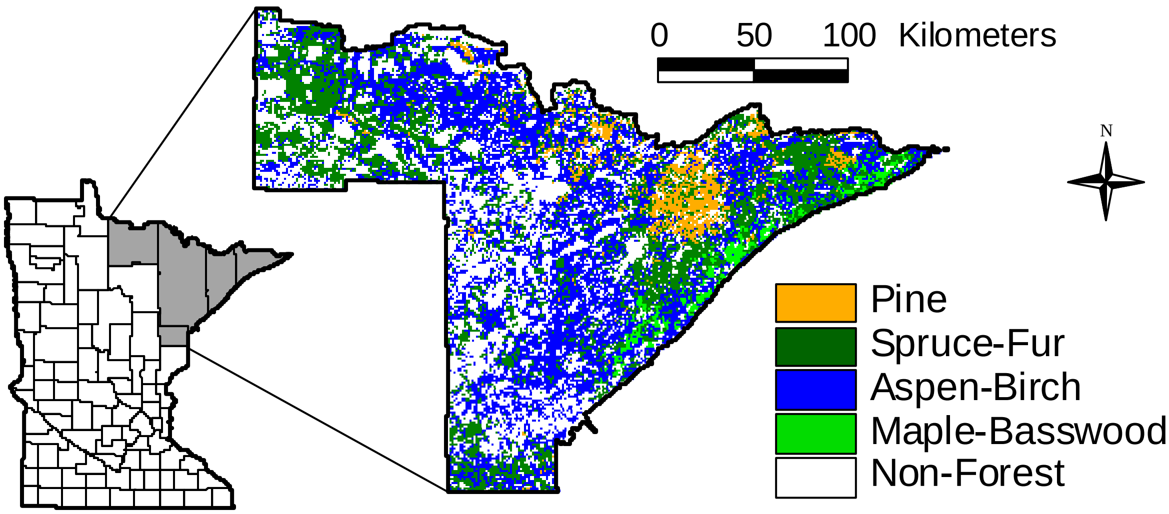


Figure 1. Study area in northeastern Minnesota.

<u>Models</u>

Process-based forest production models that are applicable at regional scales (i.e., on the order of 10^2 to 10^5 km^2) model the carbon (C) cycle by focusing on the processes of photosynthesis, respiration, and allocation of C within trees or a forest stand, and are alternatively called Ecosystem Process Models. The amount of photosynthesis and respiration is usually calculated with a "radiation-use efficiency" approach, in which total potential photosynthesis is determined and then reduced based on any number of environmental modifiers such as vapor pressure deficit and water availability in the soil. Such models typically view a forest as a single homogenous unit (i.e., "Big Leaf" model) instead of as a set of individual trees.

The two models examined in this project, PnET-II and 3-PG, are generalized, Big-Leaf type, radiation-use efficiency models intended to be applied at stand to regional scales. They were selected because of the widespread interest in them, their focus on generalized relationships and parameterizations, and the relatively few data input requirements. PnET-II (Photosynthesis and Evapotranspiration) is a lumped-parameter model of C and water balances that combines process-based and empirical components. PnET-II runs on a monthly-time step and has no specific spatial dimension, although it is commonly applied at small watershed to regional scales. For a detailed description, see Aber and Federer (1992), Aber et al. (1993), Aber et al. (1995), and Ollinger et al. (1998). 3-PG (Physiological Processes Predicting Growth) is a newer ecosystem model receiving considerable attention in the forest modeling community. Developed by Landsberg and Waring (1997), 3-PG was designed for use as both an operational and a research tool. As with PnET-II, 3-PG is a generalized, monthly time-step, radiation-use efficiency model of forest growth that includes both process-based and empirical components. A key difference with 3-PG is the inclusion of allometric relationships, where biomass outputs are converted into units important to forest managers (e.g., stand density and volume). For a detailed description, see Landsberg and Waring (1997), Coops et al. (1998), and Landsberg et al. (2002). Law et al. (2001) conducted a direct comparison of 3-PG and PnET-II, and concluded that the models were comparable and applicable for estimating annual production at the stand level.

Data

Both spatial and aspatial data were required for this project. All of the data, with the exception of model parameters, were collected from publicly available data sources and organized within a GIS. The required data can be grouped into four general categories: forest inventory data, climate variables, site variables, and model parameters. Baseline and validation forest type and inventory data were collected from the USFS FIA database for the years between the 4th and 5th Minnesota FIA surveys, 1977-1990 (Jakes 1980; Leatherberry et al. 1995). Long-term normal climate data, including monthly temperature, precipitation and solar radiation, were collected for climate stations in the region (Baker et al. 1985; Minnesota Climatology Working Group 2003; Maxwell et al. 1995). For temperature and precipitation, monthly raster grids were interpolated using an Inverse Distance-Weighted (IDW) algorithm with squared exponential weighting. For solar radiation, monthly raster grids were linearly interpreted along a latitudinal gradient from the two sample locations in the study area where data were publicly available (Duluth and International Falls, MN). The two required soil variables, soil water holding capacity (SWHC) and soil texture type, were extracted from the nation-wide STATSGO soil data set (SCS 1991).

Process-based ecosystem models use parameter values to distinguish the characteristics and responses to stimuli of different vegetation types. In this project, four general forest types were selected and parameterized: Aspen-Birch (39% of the study area), Spruce-Fir (26%), Pine (5%), and Maple-Basswood (2%). The remainder of the study area (28%) consists of non-forest areas, including open water, wetlands, developed areas, and bare land. The models were not run for these non-forest areas. For PnET-II, parameter values existed from other studies in the region (Reich et al, 1999; Sellers et al., 1997). For 3-PG, individual parameter values were extracted from published studies where possible, or else the default values were used.

Analysis

Two primary lattice grids (in the form of point feature vector files) were created for the study area, one for each of the primary spatial scales of analysis (1x1 km and 10x10 km cells). Each model was run for each forest grid cell in the study area, resulting in 25,885 runs of each model for the 1x1 km scale and 299 runs each for the 10x10 km scale. Although both models provide a variety of outputs, only aboveground Net Primary Production (NPP, or the annual sum of foliage and stem biomass production) growth estimates are evaluated here. This limited analysis was chosen because forest managers are typically more interested in above ground growth (particularly stem growth) than below ground growth. Due to their widespread applicability, the two models are treated as “black boxes” and are assumed to accurately portray regional forest production processes.

Three primary sets of analyses were conducted. First, modeled outputs were compared between the two models and across the two spatial scales using non-parametric statistical tests. Second, spatial patterns between models and across scales were compared using the Moran's I index of spatial autocorrelation. Moran's I quantifies spatial autocorrelation by using a weighted correlation coefficient to test for departures from spatial randomness (Cliff and Ord 1981). The value for Moran's I falls between -1 and 1, where 0 equals spatial randomness and negative or positive numbers indicate the level of negative or positive autocorrelation, respectively. For this project, Moran's I was calculated using the S-Plus for ArcView 3.x extension (S-Plus, 1998). The final set of analyses consisted of comparing modeled outputs against other forest production estimates in the region and the FIA database growth estimates. Additional analyses were conducted, including variogram and Latin Hypercube Sampling sensitivity analysis, but are presented in full elsewhere (Kirk 2004).

Framework

The ultimate objective of this study was to identify the core components of implementing process-based models and develop a framework for incorporation of those components. The framework developed here is explained in detail in the discussion section below.

RESULTS

The mean NPP estimates for the four modeling sets (i.e., two models run across two spatial scales) range from 783.2 to 820.9 g C m^{-2} yr^{-1}, with standard deviations between 218.7 and 255.0 g C m-2 yr^{-1}. Thus, in the most general sense, all model predictions fall in a comparable range. Figure 2 contains histograms of the NPP estimates. In aggregate, the data are multi-modal at both resolutions with no discernable patterns. This is explained by cover type differences, for which the models are parameterized. Separating by cover type results in unimodal distributions for all cover types except for the 10x10 km Aspen-Birch group, which is bimodal. Overall, none of the cover type predictions are normally distributed (chi-square test for normality, $p < 0.001$). As a result, non-parametric tests are used for comparisons presented below.

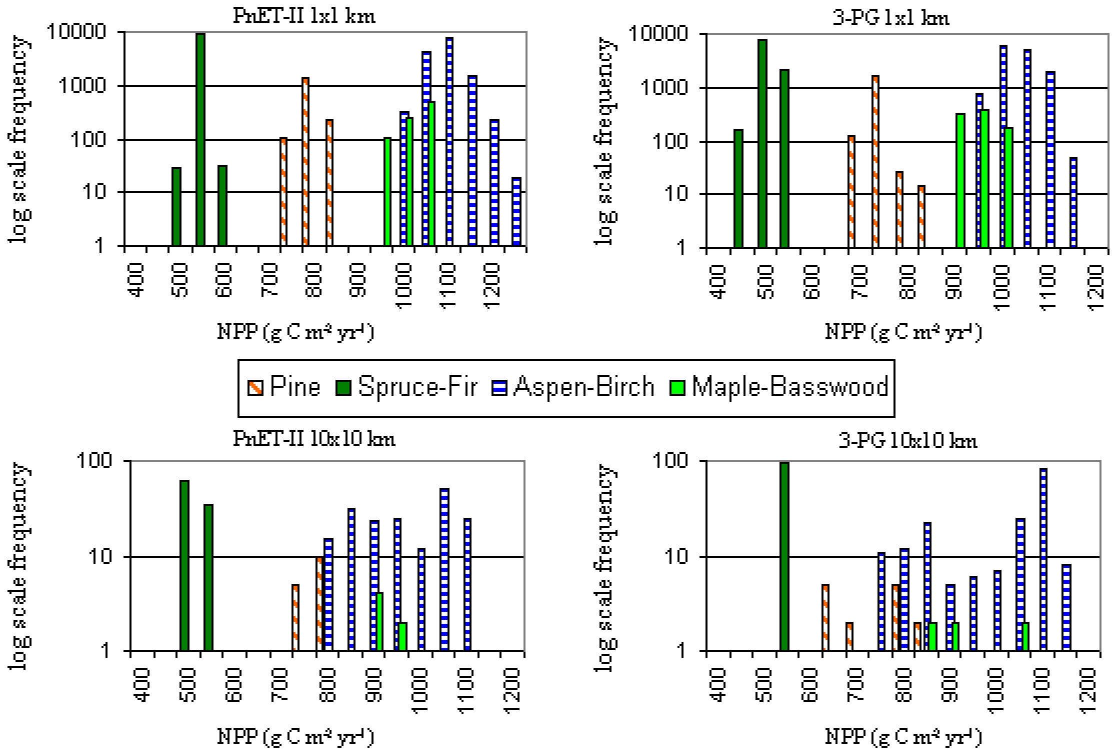


Figure 2. Log-scale histograms of the 1x1 km resolution model predictions for PnET-II and 3-PG.

In general, the model predictions are in relative agreement (Figure 3). However, there are some notable differences. At the finer spatial resolution (1x1 km), PnET-II predictions are significantly higher than the 3-PG estimates for all cover types (Wilcoxon Rank-Sum Test, $p<0.001$). For the 10x10 km resolution, however, there are no systematic patterns across all cover types. The scatterplots also illustrate the relative variance in estimates between the two models. For the equilibrium PnET-II model, the variation within each cover type is relatively low, while the non-equilibrium 3-PG estimates vary over a much wider range.

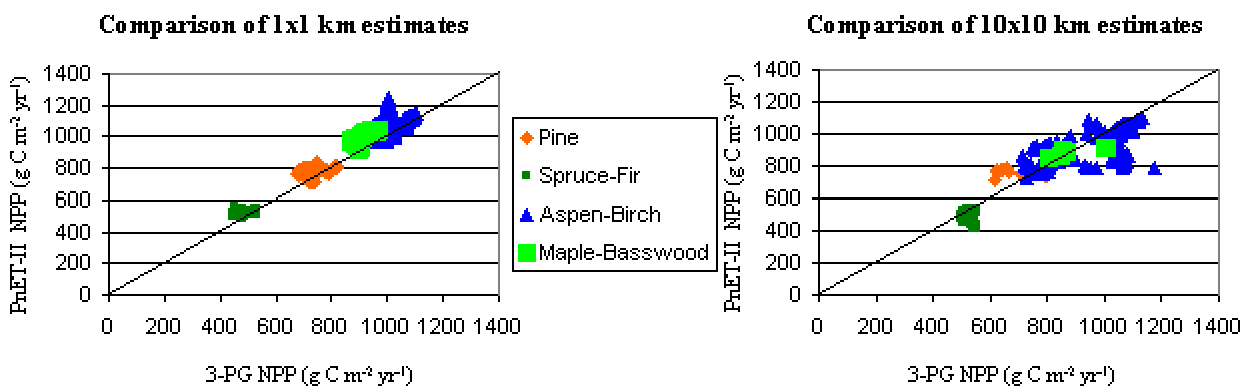


Figure 3. Scatterplot diagrams comparing estimates from the two models.

On a per-meter basis, increasing spatial resolution has a significant effect on model predictions (Wilcoxon Rank-Sum Test, $p < 0.001$). That is, mean predictions were significantly different between the 1x1 km and 10x10 km predictions for both models and all cover types (Table 1).

However, there were no systematic changes in the patterns of variation within and between cover types. For example, aggregating the Spruce-Fir cover type led to an increase in variation for the PnET-II predictions, but a decrease in variation for the 3-PG predictions. The reverse pattern is true for the Maple-Basswood cover type. This suggests that aggregation may have non-uniform and non-linear effects between the mean and variance. Thus, it appears unsafe to assume that aggregating data will have a uniform and linear influence on model predictions.

Table 1: Scale effects on average model NPP predictions.

	Cover Type									
	Overall		Pine		Spruce-Fir		Aspen-Birch		Maple-Basswood	
Resolution	1x1 km	10x10 km	1x1 km	10x10 km	1x1 km	10x10 km	1x1 km	10x10 km	1x1 km	10x10 km
Num. Cells	25885	299	1720	15	9599	97	13658	181	908	4
% Area	100	100	6.6	5	37.1	32.4	52.8	60.5	3.5	1.3
PnET-II Mean (Std. Dev)	842.8 (255.0)	783.2 (218.7)	783.2 (17.9)	757.9 (19.3)	524.9 (9.4)	494.3 (19.2)	1064.0 (33.6)	936.6 (99.3)	989.3 (55.1)	889.0 (25.3)
3-PG Mean (Std. Dev)	788.8 (246.6)	820.9 (233.8)	729.1 (18.9)	713.5 (77.0)	482.7 (20.8)	527.5 (9.26)	1003.4 (37.7)	984.5 (123.7)	911.1 (44.1)	898.1 (85.8)

Figure 4 provides maps of the NPP predictions for the two models and two spatial scales. From visual observation it appears that the data are spatially autocorrelated. Some general trends are discernable. For example, the area of highest productivity is located in the south central portion of the study area that contains the more productive upland forests.

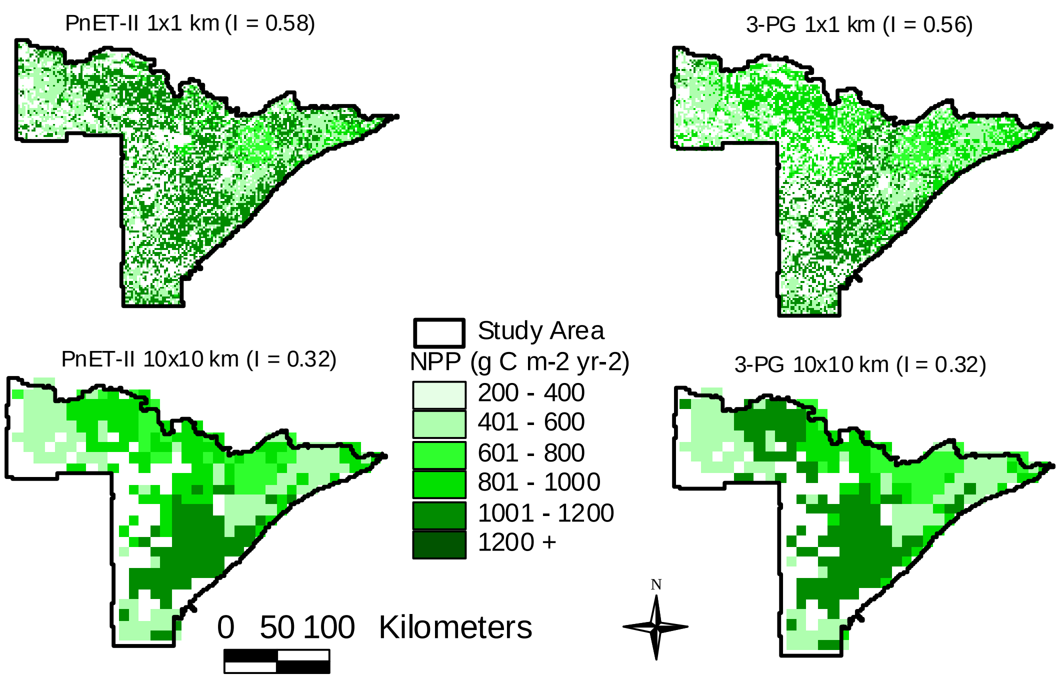


Figure 4. NPP predictions for the two models and two spatial resolutions, including Moran's I values.

Moran's I values indicate a positive spatial autocorrelation for all four modeling scenarios, with the two 1x1 km scenarios having a similar and higher autocorrelation than the 10x10 scenarios (Figure 4). The positive autocorrelation matches a visual interpretation of the NPP maps. This decreased autocorrelation at the coarser scale could be due to the large perimeter and relatively high number of no-value cells for the coarse data.

Model validation was not performed for any specific location within the study area. Because our objectives focus on regional forest production estimates, we chose to only compare modeled estimates against those from other studies in the region (Fassnacht and Gower 1996; Hall et al. 1992) and against aggregated estimates from the FIA database (Brown and Schroeder 1999). Table 2 compares NPP predictions for previous studies within or near our study area.

Table 2: Comparison of published NPP estimates against estimates from this study.

		NPP Estimate Range (g C m^{-2} yr^{-1})		
Cover Type	Source	External	PnET-II	3-PG
Pine	Fassnacht & Gower, 1996	390 – 850	707 – 836	615 – 819
Maple-Basswood	Fassnacht & Gower, 1996	290 – 1150	768 – 1047	728 – 1011
Spruce-Fir	Hall et al., 1992	40 – 572	425 – 561	444 – 549
Aspen-Birch	Hall et al., 1992	190 – 1199	724 – 1250	713 – 1174
Hardwood species	Brown & Schroeder, 1999	410 – 800	724 – 1250	713 – 1174
Softwood Species	Brown & Schroeder, 1999	210 – 600	425 – 836	444 – 819

In all cases but one, both the PnET and 3-PG estimates fell below the upper range of estimates from the other studies. However, in all cases, the minimum estimates from the other studies were well below the minimum estimates from our study. Thus, the predictions in this study may have an upward bias on NPP estimates. The sharpest contrast in estimates is between our study and the Hardwood Species estimates from the FIA based study (Brown and Schroeder 1999). Brown and Schroeder use a county level aggregation and a slightly different definition of Aboveground NPP (i.e., focusing on woody biomass), which may account for part of the difference in estimates.

DISCUSSION

This study highlights several key considerations in process-based models of forest production. On the positive side, obtaining reasonable forest production estimates is possible using readily available software tools and data sets, and once a modeling framework is established, a vast variety of analyses can be conducted. On the negative side, this study suggests that scale effects, data aggregation, and parameterization (among other things) can significantly influence or bias model predictions and results. Clearly, the decisions made during model implementation are very important. As such, a well-organized and thorough implementation framework is paramount to the success of model application.

Forest growth models have often been implemented for independent, single-use studies. Although data and the core model are regularly reused, the modeling process flow, technological setup and analyses are often recreated for each separate modeling study. There are, however, several identifiable model implementation components that are common across studies. For example, studies typically contain data preparation, parameter calibration, sensitivity analysis and report resulting phases.

For this framework (Figure 5), we identify four general components of model implementation. First, there is the model itself, which is considered here to be a "black box" model for which the inputs and outputs are the primary concern, not the internal model algorithms. Second, model implementations require data sources and data stores for both inputs and outputs. Third, technological processing is required in order to manage data and convert between formats. Finally, there are ecological model considerations, which include the scientifically based methods and decisions generally included with model studies.

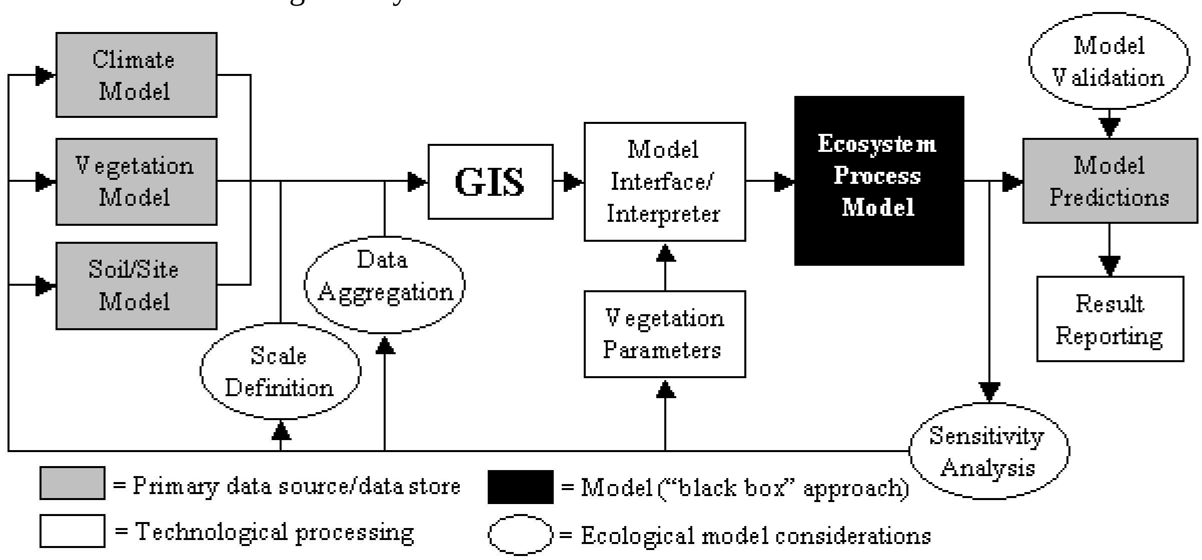


Figure 5. Conceptual framework for implementing regional scale process-based forest production models.

A fundamental decision in spatially explicit, process-based modeling is the relationship between the model and the GIS. Figure 5 contains a conceptual flowchart for a traditional "model-centered" framework. This approach requires the user to incorporate the various components, understand and properly address the ecological model considerations, and properly validate and report the results, and is arguably the most common approach in scientific modeling studies. Mitasova and Mitas (2002) identify this approach as an "import/export" (or "loose coupling") interface between the model and GIS (Figure 6). Two alternatives include an "embedded coupling," or full integration of the model within the GIS, and a "tight coupling," which provide of a common interface between the GIS and model.

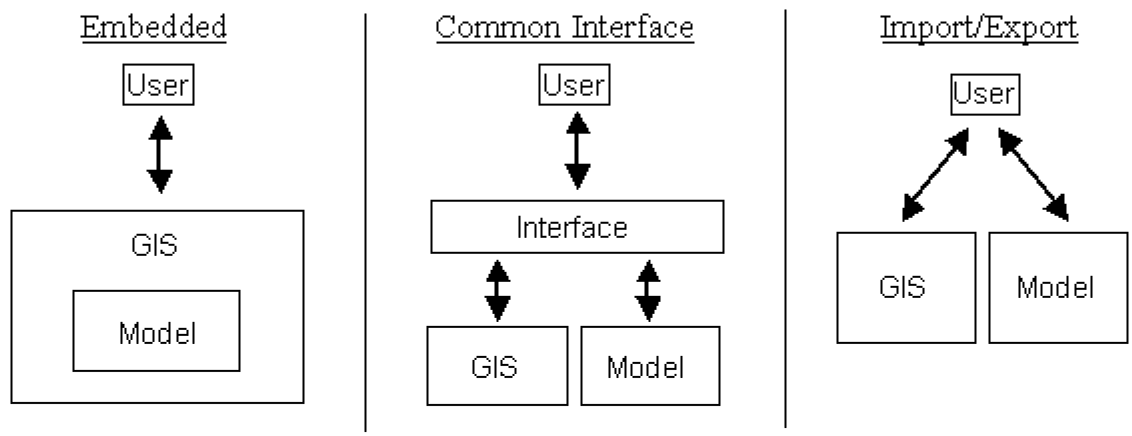


Figure 6. Visualizations of three categories of integration between models and GIS.

Overall, there is a trade-off between end-user friendliness of the more fully integrated alternatives and development flexibility of the less integrated alternatives. With recent advances in GIS software capabilities, such as the Microsoft COM-compliant structure in both ESRI's ArcGIS software suite and IDRISI's Kilimanjaro package, it is becoming easier to integrate models within a GIS. Conversely, popular scripting languages such as Perl and Python can be used as relatively simple and flexible "wrapper" programs in which to link the various model components in a less rigid arrangement. Thus, the technological tools are available to meet a variety of modeling approaches.

Another decision point in process-based modeling relates to topics primarily in the scientific domain, including parameterizations, sensitivity analyses, scale effects, and validation. Over the past several years university-level courses, public short courses, and publications have been developed to explore the alternatives and consequences of these modeling considerations. Additionally, a variety of software tools have been developed to aid the process. For example, PEST (S.S. Papadopulos and Associates, Inc. 2004) is a model-independent parameter estimation and uncertainty analysis package used across many scientific and engineering disciplines. Clearly, a wide assortment of resources are available to aid the model implementation process.

Finally, this study supports the argument that every study is only as good as its data. The results of this study suggest that it is possible to get reasonable estimates of forest production with generic data. With improvements in data quality, model prediction accuracy will improve. Acquisition and assimilation of various data sets into a common format, resolution and extent is challenging. As with most types of GIS analyses, special attention should be given to the data preparation phase.

CONCLUSIONS

This study used a sample modeling implementation in northeastern Minnesota as the basis for identifying a framework for implementing regional-scale forest production models. Using the PnET-II and 3-PG models with publicly available data sets, forest production estimates were comparable to estimates from other studies in the region in terms of mean production estimates and spatial patterns. The four general components of the modeling framework include the model itself, technological processing steps, data sources and stores, and ecological modeling considerations.

It is apparent that specialized training is required to implement process-based models in a production environment. This includes general understanding of the ecological processes being modeled, specific knowledge of the computer-technologies being used, and scientific comprehension of the statistical, analytical, and scaling issues surrounding process-based models. With improved model implementations and proper training, a myriad of new management modeling analyses are possible.

To date, there is only one documented example of the 3-PG model being used in an operational environment (Almeida et al. 2002), and no known operational uses of PnET-II have been identified. Indeed, process-based models have yet to find a widespread niche in traditional forest management. While citing several key challenges for using process-based models operationally (e.g., improving C allocation routines, scaling, etc), Landsberg (2003) identifies the primary challenge as a need for communication between model developers and forest managers. Addressing the information system and GIS components may provide a common ground for beginning this communication.

Acknowledgements
Research supported by University of Minnesota Agricultural Experiment Station Project 042-44.

REFERENCES

1. Aber, J. D., and C. A. Federer. 1992. A generalized, lumped-parameter model of photosynthesis, evapotranspiration and net primary production in temperate and boreal forest ecosystems. *Oecologia* 92: 463-474.
2. Aber, J. D., C. T. Driscoll, C. A. Federer, R. Lathrop, G. Lovett, J. M. Melillo, P. Steudler and J. Vogelmann. 1993. A strategy for the regional analysis of the effects of physical and chemical climate change on biogeochemical cycles in northeastern (U.S.) forests. *Ecological Modeling* 67:37-47.
3. Aber, J. D., S. V. Ollinger, C. A. Federer, P. B. Reich, M. L. Goulden, D. W. Kicklighter, J. M. Melillo and R. G. Lathrop, Jr. 1995. Predicting the effects of climate change on water yield and forest production in the Northeastern U.S. *Climate Research* 5: 207-222.
4. Almeida, A.C., Landsberg, J.J., Sands, P.J., 2004a. Parameterisation of 3-PG model for fast growing Eucalyptus grandis plantations. *Forest Ecology and Management* 193, 179-196.
5. Baker, D.G., E.L. Kuehnast, and J.A. Zandlo. 1985. Climate of Minnesota: Part XV: Normal Temperatures (1951-1980) and Their Application. University of Minnesota Agricultural Experiment Station paper AD-SB-2777.
6. Battaglia, M. and P. J. Sands. 1998. Process-based forest productivity models and their application in forest management. *Forest Ecology and Management* 102: 13-32.
7. Brown, S. L. and P. E. Schroeder. 1999. Spatial patterns of aboveground production and mortality of woody biomass for eastern U.S. Forests. *Ecological Applications* 9(3): 968-980.

8. Cliff, A. D. and J. K. Ord. 1981. *Spatial processes: Models and applications*. Pion Ltd., London.

9. Coops, N. C., R. H. Wharing, and J. J. Landsberg. 1998. Assessing forest productivity in Australia and New Zealand using a physiologically-based model driven with averaged monthly weather data and satellite-derived estimates of canopy photosynthetic capacity. *Forest Ecology and Management* 104: 113-127.

10. Fassnacht, K. S. and S.T. Gower. 1997. Interrelationships among the edaphic and stand characteristics, leaf area index, and aboveground net primary production of upland forest ecosystems in north central Wisconsin. *Canadian Journal of Forest Research* 27: 1058-1067.

11. Hall, F. G., K. F. Huemmrich, D. E. Strebel, S. J. Goetz, J. E. Nickeson, and K. D. Woods. 1996. Forest Biophysical Parameters (SNF). Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee

12. Jakes, P. J. 1980. The fourth Minnesota forest inventory: area. USDA Forest Service North Central Forest Experiment Station Resource Bulletin NC-54.

13. Johnsen, K., L. Samuelson, R. Teskey, S. McNulty, and T. Fox. 2001. Process models as tools in forestry research and management. *Forest Science* 47: 2-8.

14. Kirk, R. K. 2004. A technical analysis of regional-scale forest production modeling in Northeastern Minnesota. M.S. project. University of Minnesota, St. Paul, MN.

15. Korzukhin, M. D., Ter-Mikaelian, M. T. and R. G. Wagner. 1996. Process versus empirical models: which approach for forest ecosystem management? *Canadian Journal of Forest Research* 26: 879-887.

16. Landsberg, J. J. 1986. *Physiological Ecology of Forest Production*. London: Academic Press.

17. Landsberg, J. J. and S. T. Gower. 1997. *Applications of Physiological Ecology to Forest Management.* San Diego: Academic Press.

18. Landsberg, J. J. and Waring, R. H. 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.

19. Landsberg, J. J., R. H. Waring, and N. C. Coops. 2002. Performance of the forest productivity model 3-PG applied to a wide range of forest types. *Forest Ecology and Management* 172: 199-214.

20. Landsberg, J. J. 2003. Modelling forest ecosystems: state of the art, challenges, and future directions. *Canadian Journal of Forest Research* 33: 385-397.

21. Law, B. E., R. H. Waring, P. M. Anthoni, and J. D. Aber. 2000. Measurement of gross and net ecosystem productivity and water vapor exchange of a Pinus ponderosa ecosystem, and an evaluation of two generalized models. *Global Change Biology* 6:155-168.

22. Leatherberry, E. C., J. S. Spencer, T. L. Schmidt, and M. R. Carroll. 1995. An analysis of Minnesota's fifth forest resources inventory, 1990. USDA Forest Service North Central Forest Experiment Station Resource Bulletin NC-165.

23. Mäkelä, A., J. Landsberg, A. R. Ek, T. E. Burk, M. Ter-Mikaelian, G. I. Ågren, C. D. Oliver and P. Puttonen. Process-based models for forest ecosystem management: current state of the art and challenges for practical implementation. *Tree Physiology* 20:289-298.

24. Maxwell, E. L., W. Marion, D. Myers, M. Rymes, and S. Wilcox. 1995. *Final technical report: National Solar Radiation Database (1961-1990).* National Renewable Energy Laboratory, Golden, CO, NREL/TP-463-5784.

25. Minnesota Climatology Working Group. 2003. Available at: http://climate.umn.edu/doc/historical/normals.htm. Accessed 16 August 2004.

26. Mitasova, H. and L. Mitas. 2002. Modeling Physical Systems. In: Clarke, K.C., B.O. Parks, and M.P. Crane, eds. *Geographic Information Systems and Environmental Modeling*. Prentice Hall: Upper Saddle River.

27. Mohren, G. M. J., and H. E. Burkhart. 1994. Contrasts between biologically-based process models and management-oriented growth and yield models. *Forest Ecology and Management* 69: 1-5.

28. Ollinger, S. V. J. D. Aber, and C. A. Federer. 1998. Estimating regional forest productivity and water yield using an ecosystem model linked to a GIS. *Landscape Ecology* 13: 323-334.

29. PEST. 2004. *PEST for Windows*. Ver. 8.0. Boulder, CO. S.S. Papadopulos & Associates, Inc.

30. Reich, P. B., D. P. Turner, and P. V. Bolstad. 1999. An approach to spatially distributed modeling of net primary production (NPP) at the landscape scale and its application in validation of EOS NPP products. *Remote Sensing of Environment* 70: 69-81.

31. SCS. 1991. State soil geographic database (STATSGO) data users guide. U.S. Soil Conservation Service Miscellaneous Publication 1492.

32. Sellers, P. J., F. G. Hall, R. D. Kelly, et al. 1997. BOREAS in 1997: Experiment Overview, Scientific Results and Future Directions. *Journal of Geophysical Researc*h 102: 731-770.

33. Sievänen, R., and T. E. Burk. 1993. Adjusting a process-based growth model for varying site conditions through parameter estimation. *Canadian Journal of Forest Research* 23: 1837-1851.

Literature Cited

Aber, J. D., and C. A. Federer. 1992. A generalized, lumped-parameter model of photosynthesis, evapotranspiration and net primary production in temperate and boreal forest ecosystems. *Oecologia.* 92: 463-474.

Aber, J. D., C. T. Driscoll, C. A. Federer, R. Lathrop, G. Lovett, J. M. Melillo, P. Steudler and J. Vogelmann. 1993. A strategy for the regional analysis of the effects of physical and chemical climate change on biogeochemical cycles in northeastern (U.S.) forests. *Ecological Modeling* 67:37-47

Aber, J. D., S. V. Ollinger, C. A. Federer, P. B. Reich, M. L. Goulden, D. W. Kicklighter, J. M. Melillo and R. G. Lathrop, Jr. 1995. Predicting the effects of climate change on water yield and forest production in the Northeastern U.S.

Aber, J. D., P. B. Reich and M.I. Goulden. 1996. Extrapolating leaf CO2 exchange to the canopy: a generalized model of forest photosynthesis validated by eddy correlation. *Oecologia*. 106: 257-265.

Aber, J. D., and C. T. Driscoll. 1997. Effects of land use, climate variation, and N deposition on N cycling and C storage in the northern hardwood forests. *Global Biogeochemical Cycling*. 11”639-48.

Aber, J. D., and J. M. Melillo. 1991. *Terrestrial Ecosystems*. Saunders: Philadelphia. 429 p.

Almeida, A. C., R. Maestri, J. J. Landsberg, and J. R. S. Scolforo. 2002. Linking process-based and empirical forest models to use as a practical tool for decision-making in fast growing Eucalyptus plantation in Brazil. Abstracts from the IUFRO conference on Reality, models and parameter estimation – the forestry scenario. 2-5 June 2002, Sesimbra, Portugal.

Baker, D.G., E.L. Kuehnast, and J.A. Zandlo. 1985. Climate of Minnesota: Part XV: Normal Temperatures (1951-1980) and Their Application. University of Minnesota Agricultural Experiment Station paper AD-SB-2777.

Baskent, E. Z. ,and H. A. Yolasigmaz. 1999. Forest Landscape Management Revisited. *Environmental Management*. 24: 437-448.

Bishop, G. D., M. R. Church, J. D. Aber, R. P Neilson, S. V. Ollinger, and C. Daly. 1998. A comparison of mapped estimates of long-term runoff in the northeast United States. *Journal of Hydrology*. 206: 176-190.

Battaglia, M. and P. J. Sands. 1998. Process-based forest productivity models and their application in forest management. *Forest Ecology and Management.* 102: 13-32.

Bolstad, P.V., L. Swift, F. Collins and J. Regniere. 1998. Measured and predicted air temperatures at basin to regional scales in the Southern Appalachian Mountains. *Agric. For. Meteorol.* 91:167--176.

Bolstad, P. V. 2002. *GIS Fundamentals: A First Text on Geographic Information Systems*. White Bear Lake: Eider Press. 412 p.

Brown, S. L. and P. E. Schroeder. 1999. Spatial patterns of aboveground production and mortality of woody biomass for eastern U.S. Forests. *Ecological Applications*. 9(3): 968-980.

Campbell, J. B. 1996. *Introduction to Remote Sensing*. New York: Guilford Press. 622 p.

Coops, N. C., R. H. Wharing, and J. J. Landsberg. 1998. Assessing forest productivity in Australia and New Zealand using a physiologically-based model driven with averaged monthly weather data and satellite-derived estimates of canopy photosynthetic capacity. *Forest Ecology and Management*. 104: 113-127.

Coops, N. C., and R. H. Waring. 2001a. Estimating forest productivity in the eastern Siskiyou Mountains of southwestern Oregon using a satellite driven process model, 3-PGS. *Canadian Journal of Forest Research*. 31: 143-154.

Coops, N. C., and R. H. Waring. 2001b. Assessing forest growth across southwestern Oregon under a range of current and future global change scenarios using a process model, 3-PG. *Global Change Biology*. 7: 15-29.

Coops, N. C., and R. H. Waring. 2001c. The use of multiscale remote sensing imagery to derive regional estimates of forest growth capacity using 3-PGS. *Remote Sensing of Environment*. 75: 324-334.

Dye, P. J. 2001. Modelling growth and water use in four *Pinus patula* stands with the 3-PG model. *Southern African Forestry Journal*. 191: 53-

Fassnacht, K. S. and S.T. Gower. 1997. Interrelationships among the edaphic and stand characteristics, leaf area index, and aboveground net primary production of upland forest ecosystems in north central Wisconsin. *Canadian Journal of Forest Research*. 27: 1058-1067.

Field, C. B. and J. R. Ehleringer (Eds.). 1993. *Scaling Physiological Processes: Leaf to Globe.* San Diego: Academic Press. 388 p.

Field, C. and H. A. Mooney. 1986. The photosynthesis-nitrogen relationships in wild plants. In: Givnish, T. ed. *On the economy of plant form and function.* Cambridge: Cambridge University Press. P 25-55.

Gbondo-Tugbawa, S.S., C.T. Driscoll , J.D. Aber and G.E. Likens. The evaluation of an integrated biogeochemical model (PnET-BGC) at a northern hardwood forest ecosystem. *Water Resources Research* (in press)

Goodale, C. L., J. D. Aber and E. P. Farrell. 1998. Predicting the relative sensitivity of forest production in Ireland to site quality and climate change. *Climate Research.* 10:51-67

Gower, S. T., C. J. Kucharik, J. M. Norman. 1999. Direct and indirect estimation of Leaf Area Index, fPAR, and Net Primary Production on terrestrial ecosystems. *Remote Sensing of Environment*. 70: 29-51.

Gunderson, L. H., C. S. Holling, and S. S. Light. 1995. *Barriers and bridges to the renewal of ecosystems and institutions*. New York: Columbia Press.

Hall, F. G., K. F. Huemmrich, D. E. Strebel, S. J. Goetz, J. E. Nickeson, and K. D. Woods. 1996. Forest Biophysical Parameters (SNF). [Forest Biophysical Parameters (Superior National Forest)]. Data set. Available on-line [http://www.daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. Based on F. G. Hall, K. F. Huemmrich, D. E. Strebel, S. J. Goetz, J. E. Nickeson, and K. D. Woods, Biophysical, Morphological, Canopy Optical Property, and Productivity Data from the Superior National Forest, NASA Technical Memorandum 104568, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, U.S.A., 1992.

Hansen, M. H., T. Frieswyk, J. F. Glover, and J. F. Kelly. 1992. The Eastwide forest inventory data base: user manual. USDA Forest Service North Central Forest Experiment Station General Technical Report NC-151.

Henderson-Sellers, B., and A. Henderson-Sellers, 1996: Sensitivity evaluation of environmental models using fractional factorial analysis. *Ecol. Modelling*, 86: 291–295.

He, H. S., D. J. Mladenoff, V. C. Radeloff, and T. R. Crow. 1998. Integration of GIS data and classified satellite imagery for regional forest assessment. *Ecological Applications*. 8: 1072-1083.

Hungeford, R. D., R. R. Nemani, S. W. Running, and J. C. Coughlan. 1989. MTCLIM: A mountain microclimate simulation model. USDA Forest Service Intermountain Research Station Research Paper INT-414.

Isaaks, E. H. and R. M. Srivastava. 1989. *An Introduction to Applied Geostatistics*. New York: Oxford Press. 561 p.

Isard, W. 1975. *An Introduction to Regional Science*. Englewood Cliffs: Prentice-Hall. 506. pl.

Jakes, P. J. 1980. The fourth Minnesota forest inventory: area. USDA Forest Service North Central Forest Experiment Station Resource Bulletin NC-54.

Jarvis, P. G. 1995. Scaling processes and problems. *Plant, Cell and Environment*. 18: 1079-1089.

Jenkins, J. C., D. W. Kicklighter, S. V. Ollinger, J. D. Aber, and J. M. Melillo. 1999. Sources of variability in net primary production predictions at a regional scale: a comparison using PnET-II and TEM 4.0 in northeaster U.S. forests. *Ecosystems*. 2: 555-570.

Johnsen, K., L. Samuelson, R. Teskey, S. McNulty, and T. Fox. 2000. Process models as tools in forestry research and management. *Forest Science*. 47: 2-8.

Kingsley, N. P. 1991. Forest statistics for Minnesota's Aspen-Birch Unit. USDA Forest Service North Central Forest Experiment Station Resource Bulletin NC-128.

Kirk, Ryan W., and Thomas E. Burk. 2004. Regional-Scale forest production modeling using process-based models and GIS. In: SoFor GIS 2004: Proceedings of the 4th Southern Forest and Natural Resources Geographic Information Systems Conference. Athens, GA, December 2004.

Korzukhin, M. D., Ter-Mikaelian, M. T. and R. G. Wagner. 1996. Process versus empirical models: which approach for forest ecosystem management? *Canadian Journal of Forest Research*. 26: 879-887.

Landsberg, J. J. 1986. *Physiological Ecology of Forest Production*. London: Academic Press 198 p.

Landsberg, J. J. and S.T. Gower 1997. *Applications of Physiological Ecology to Forest Management.* San Diego: Academic Press. 354 p.

Landsberg, J. J. 2001. Applying 3-PG, a simple process-based model designed to produce practical results, to data from loblolly pine experiements. *Forest Science*. 47: 43-51.

Landsberg, J. J., R. H. Waring, and N. C. Coops. 2002. Performance of the forest productivity model 3-PG applied to a wide range of forest types. *Forest Ecology and Management*. In press.

Lathrop, R. G., J. D. Aber, and J. A. Bognar. 1995. Spatial variability of a digital soils map in a regional modeling context. *Ecological Modeling*. 82: 1-10.

Leatherberry, E. C., J. S. Spencer, T. L. Schmidt, and M. R. Carroll. 1995. An analysis of Minnesota's fifth forest resources inventory, 1990. USDA Forest Service North Central Forest Experiment Station Resource Bulletin NC-165.

Liu, J., C. Peng, Q. Dang, M. Apps, and H. Jiang. 2002. A component object model strategy for reusing ecosystem models. *Computers and Electronics in Agriculture*. 35: 17-33.

Lucas, N.J., and P.J. Curran. 1999. Forest ecosystem simulation modelling: the role of remote sensing. *Progress in Physical Geography*. 23(3): 391-423.

Mäkelä, A., J. Landsberg, A. R. Ek, T. E. Burk, M. Ter-Mikaelian, G. I. Ågren, C. D. Oliver and P. Puttonen. Process-based models for forest ecosystem management: current state of the art and challenges for practical implementation. *Tree Physiology.* 20:289-298.

Maxwell, E. L., W. Marion, D. Myers, M. Rymes, and S. Wilcox. 1995. *Final technical report: National Solar Radiation Database (1961-1990).* National Renewable Energy Laboratory, Golden, CO, NREL/TP-463-5784.

McKay, M. D., W. J. Conover, and R. K. Beckman. 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21: 239-45.

McNulty, S. G., J. M. Vose, W. T. Swank, J. D. Aber, and C. A. Federer. 1994. Regional-scale forest ecosystem modeling: database development, model predictions and validation using a geographic information system. *Climate Research*. 11: 109-124.

McNulty, S. G., J. M. Vose, and W. T. Swank. 1996. Potential climate change effects on loblolly pine forest productivity and drainage across the Southern United States. *Ambio*. 25: 449-453.

McNulty, S. G., J. M. Vose, and W. T. Swank.1997. Regional hydrologic response of loblolly pine to air temperature and precipitation changes. *Journal of the American Water Resources Association*. 33: 1011-1022.

McNulty, S.G., J.M. Vose, and W. T. Swank. 1997. Scaling predicted pine forest hydrology and productivity across the Southern United States. In: Quattrochi, D. A., and M. F. Goodchild, ed. *Scale in Remote Sensing and GIS*. Lewis: Boca Raton. p. 187-209.

McNulty, S. G., J. A. Moore, L. Iverson, A. Prasad, R. Abt, B. Smith, G. Sun, M. Gavazzi, J. Bartlett, B. Murray, R. Mickler and J. D. Aber. 2000. Application of linked regional scale growth, biogeography and economic models for southeastern United States pine forests. *World Resource Review* 12:298-320.

Meentemeyer, V. 1989. Geographical perspectives of space, time, and scale. *Landscape Ecology*. 3(3/4): 163-173.

Mickler, R. A., T. S. Earnhardt, and J. A. Moore. 2002. Regional estimation of current and future forest biomass. *Environmental Pollution*. 116: S7-S16.

Minnesota Climatology Working Group. 2003. http://climate.umn.edu/doc/historical/normals.htm

Mitasova, H. and L. Mitas. 1995. Modeling Physical Systems. In: Clarke, K.C., B.O. Parks, and M.P. Crane, eds. *Geographic Information Systems and Environmental Modeling*. Prentice Hall: Upper Saddle River.

Mohren, G. M. J., and H. E. Burkhart. 1994. Contrasts between biologically-based process models and management-oriented growth and yield models. *Forest Ecology and Management*. 69: 1-5.

Ollinger, S. V., J. D. Aber, C. A. Federer, G. B. Lovett, and J. Ellis. 1995. Modeling physical and chemical climatic variables across the northeastern U.S. for a Geographic Information System. USDA Forest Service Northeast Research Station General Technical Report NE-191.

Ollinger, S. V. J. D. Aber, and C. A. Federer. 1998. Estimating regional forest productivity and water yield using an ecosystem model linked to a GIS. *Landscape Ecology*. 13: 323-334.

Parton, W. J., D. S. Schimel, C. V. Cole, and D. S. Ojima. 1987. Analysis of factors controlling soil organic matter levels in Great Plains grasslands. *Soil Science Society of America Journal.* 51: 1173-1179.

Parton, W. J., J. M. Scurlock, D. S. Ojima, T. G. Golmanov, R. J. Scholes, D. S. Schimel, T. Kirchner, J. Menaut, T. Seastedt, E. Gargia Moya, A. Kamnalrut, and J. I. Kinyamario. 1993. Observations and modeling of biomass and soil organic matter dynamics for the grassland biome worldwide. *Global Biogeochemical Cycles*. 7: 785-809.0

Pastor, J. and W. M. Post. 1986. Influence of climate, soil moisture, and succession on forest carbon and nitrogen cycles. *Biogeochemistry*. 2: 3-27.

Peng, C., J. Liu, Q. dang, M. J .Apps, and H. Jiang. 2002. TRIPLEX: a generic hybrid model for predicting forest growth and carbon and nitrogen dynamics. *Ecological Modelling*. 153: 109-130.

Plummer, S.E. 2000. Perspectives on combining ecological process models and remotely sensed data. *Ecological Modelling*. 129: 169-186.

Postek, K. M., C.T. Driscoll, J. D. Aber and R. C. Santore. 1995. Application of PnET-CN/CHESS to a spruce stand in Solling, Germany. *Ecological Modelling* 83:163-172.

Radtke, P. J., T.E. Burk, and P.V. Bolstad. 2002. Bayesian melding of a forest ecosystem model with correlated inputs. *Forest Science*. 48(4): 701-711.

Radtke, P. J., T.E. Burk, and P.V. Bolstad. 2001. Estimates of the distributions of forest ecosystem model inputs for deciduous forests of eastern North America. *Tree Physiology*. 21: 505-512.

Radtke, P. J. 1999. *Uncertainty Assessment, Calibration and Sensitivity Analysis of Process-Based Forest Ecosystem Computer Simulation Models.* Ph.D. Thesis. University of Minnesota, St. Paul, MN. 318 p.

Raich, J. W., E. B. Rastetter, J. M. Melillo, D. W. Kicklighter, P. A. Steudler, B. J. Peterson, A. L. Grace, B. Moore III, and C. J. Vorosmarty. 1991. Potential Net Primary Productivity in South America: application of a global model. *Ecological Applications*. 1: 399-429.

Reich, P. B., B. Kloeppel, D. S. Ellsworth, and M. B. Walters. 1995. Different photosynthesis-nitrogen relations in deciduous and evergreen coniferous tree species. *Oecologia*. 104: 24-30.

Reich, P. B., D. P. Turner, and P. V. Bolstad. 1999. An approach to spatially distributed modeling of net primary production (NPP) at the landscape scale and its application in validation of EOS NPP products. *Remote Sensing of Environment*. 70: 69-81.

Robinson, A.P. 1998. *Forest ecosystem dynamics: systematic approach to modelling in a model rich environment.* Ph.D. Thesis. University of Minnesota, St. Paul, MN. 235 p.

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

Running, S. W., R. R. Nemani, D. L. Peterson, L. E. Band, D. F. Potts, L. L. Pierce, and M. A. Spanner. 1989. Mapping regional forest evapotranspiration and photosynthesis by coupling satellite data with ecosystem simulation. *Ecology*. 70: 1090-1101.

Running, S. W., and S. T. Gower. 1991. FOREST-BGC, A general model of forest ecosystem processes for regional applications. II. Dynamic carbon allocation and nitrogen budgets. 9: 147-160.

Schlesinger, W. H. 1997. Biogeochemistry: An Analysis of Global Change, 2nd Edition, Academic Press.

SCS. 1991. State soil geographic data base (STATSGO) data users guide. U.S. Soil Conservation Service Miscellaneous Publication 1492.

Sellers, P. J., F. G. Hall, R. D. Kelly, et al. 1997. BOREAS in 1997: Experiment Overview, Scientific Results and Future Directions. Journal of Geophysical Research 102: 731-770.

Shindler, B.1998. Landscape-Level Management: It's All About Context. *Journal of Forestry*. 98: 10-14.

Sievänen, R., and T. E. Burk. 1993. Adjusting a process-based growth model for varying site conditions through parameter estimation. *Canadian Journal of Forest Research.* 23: 1837-1851.

S-Plus. 1998. ArcView GIS Link. MathSoft, Seattle, WA.

Surfer 8.0. 2003. Golden Software, Inc., Golden, Colorado.

Tickle, P. K., N. C. Coops, S. D. Hafner, and The Bago Science Team. 2001. Assessing forest productivity at local scales across a native eucalypt forest using a process model, 3PG-Spatial. *Forest Ecology and Management*. 152: 275-291.

Turner, D.P., O. Sun, C. Daly, et al. 2003. Effects of land use and fine-scale environmental heterogeneity on net ecosystem production over a temperate coniferous forest landscape. *Tellus, Series B: Chemical and Physical Meteorology*. 55(2): 657-668.

Turner, D.P., S.V. Ollinger, and John S. Kimball. 2004. Integrating remote sensing and ecosystem process models for landscape- and regional-scal analysi of the carbon cycle. *BioScience*. 54(6): 573-584.

U.S. Department of the Interior, U.S. Geological Survey, 1987 Geographic Names Information System--Data Users Guide 6; Reston, Virginia.

Walters, D. K. 1994. *Evaluation methodology of forest ecosystem change models.* Ph.D. Thesis. University of Minnesota, St. Paul, MN. 332 p.

Waring, R. H. and W. H. Schlesinger. 1985. *Forest Ecosystems: Concepts and Management.*

Waring, R. H. and S. W. Running. 1998. *Forest Ecosystems: Analysis at Multiple Scales.* San Diego: Academic Press. 370 p.

Waring R. H., J. J. Landsberg, M. Williams. 1998. Net primary production of forests: a constant fraction of gross primary production? *Tree Physiology*. 18: 129-134.

Waring, R. H. 2000. A process model analysis of environmental limitations on the growth of Sitka spruce plantations in Great Britain. *Forestry*. 73: 65-79.

White, J. D., N. C. Coops, and N. A. Scott. 2000. Estimates of New Zealand forest and scrub biomass from the 3-PG model. *Ecological Modelling.* 131: 175-190.

White, M.A., P.E. Thorton, S.W. Running, and R. R. Nemani. 2000. Parameterization and sensitivity analysis of the BIOME-BGC Terrestrial Ecosystem Model: Net Primary Production controls. *Earth Interactions*. 4(3):1-35.