Calibration for BIOME-BGC Model through Data Assimilation with Remote Sensing Data

by

Chomchid Imvitthaya

A progress report submitted in partial fulfillment of the requirements for the degree of Doctoral of Technical Science

Examination Committee: Dr. Kiyoshi Honda (Chairperson)
Dr. Surat Lertlum
Dr. Voratas Kachitvichyanukul
Dr. Nipon TangTham

Nationality: Thai
Previous Degree: Master of Science
(Human Settlement Planning Geography),
Department of Geography,
Kasetsart University, Thailand

Scholarship Donor: Royal Thai Government Fellowship

Asian Institute of Technology
School of Engineering and Technology
Thailand

Calibration for BIOME-BGC Model through Data Assimilation with Remote Sensing Data

by

Chomchid Imvitthaya

A progress report submitted in partial fulfillment of the requirements for the degree of Doctoral of Technical Science

Examination Committee:	Dr. Kiyoshi Honda (Chairperson) Dr. Surat Lertlum Dr. Voratas Kachitvichyanukul Dr. Nipon TangTham
Nationality:	Thai
Previous Degree:	Master of Science (Human Settlement Planning Geography), Department of Geography, Kasetsart University, Thailand
Scholarship Donor:	Royal Thai Government Fellowship

Asian Institute of Technology
School of Engineering and Technology
Thailand

ACKNOWLEDGEMENT

I would like to acknowledge the Biodiversity Research and Training (BRT), Grant BRT T_551002, Thailand for making possible the conduct of this research through a Doctoral Research Grant for the period October 2007 – September 2009. Professor Dr. Visut Baimai, BRT Director is highly appreciated for coming up with such a distinguished program.

I would like also to acknowledge the Thai Royal Government and Office of Higher Education Commission through Kasetsart University for their benevolence in providing a full scholarship grant for my doctoral study.

The untiring guidance and help of my research committee, namely: Assoc. Prof. Dr. Kiyoshi Honda, Dr. Surat Lertlum, Prof. Dr. Voratas Kachitvichyanukul, and Prof. Dr. Nipon TangTham are highly appreciated.

I would like to thanks the Numerical Terradynamic Simulation Group (NTSG) at the University of Montana for allowing us to access the BIOME-BGC 4.1.1 codes. I am also very grateful to Dr. Yoshifuji, Mr. Chatchai Tantasirin, and their team for field data support.

I wish also to extend my gratitude to some special people who are always there for me in spite of:

To my beloved family, mother, Ampron, and father, Prathomasak, for being there to tap my shoulders whenever I felt down towards my work, to my sister AopEau and brother, Kukit for their unselfish commitment toward my desire to finish my doctoral degree; to my aunty, Angkana and my super friend, Sakol Promsin who serve as an inspiration for me to move on against all obstacles on my way.

I thank holy Buddha for all your help.

ABSTRACT

This study presents the results of a large scale analysis of the carbon budget of teak (*Tectona grandis* Lin F.) which is an important species in tropical deciduous forests in northern Thailand. BioGeochemical Cycles (BIOME-BGC) was calibrated to estimate net primary production (NPP) using data assimilation with remote sensing data (SPOT-LAI). Genetic Algorithm (GA) was coupled with BIOME-BGC (BIOME-BGC-GA) to find optimal ecophysiological model parameters. Calibration was performed by adjusting simulated LAI to satellite LAI, and the best fit of both of them confirmed the BIOME-BGC-GA method's accuracy. The optimized model was evaluated using the NPP satellite data, and showed good improvement to the results obtained by the BIOME-BGC-GA using the default literature parameterization. This improvement was mainly because the model's optimized parameters reduced the bias by reducing the systematic underestimation of the model. These results encourage the operational application of BIOME-BGC in teak forests in tropical zones and present an enhanced method of using BIOME-BGC-GA to develop the ecophysiological parameters that are crucial for NPP simulation. NPP plays a significant role in large scale forest carbon studies.

Table of Contents

List of Tables

List of Figures

CHAPTER I

INTRODUCTION

1.1 Background

The concentration of a greenhouse gas or carbon dioxide (CO_2) in the atmosphere has increased regularly in coincidence with the great increase in global mean air temperature since the 1910s (IPCC 2001). Due to these phenomena, decreasing the amount of the atmospheric CO_2 and understanding the carbon storage and balance in the terrestrial ecosystems are in urgent need, especially forest ecosystems. Forest ecosystems are an environmental and economic resource which is still widely spread. Forests are able to provide numerous wood products (e.g. timber, paper products, etc.), prevent soil erosion, contribute to maintain biodiversity and are often used for recreational purposes; additionally they have a great role both in the water and carbon cycles (Waring and Running, 1998). Hence, the necessity for monitoring and quantifying the amount of carbon accumulated within forests has recently increased also in view of the application of the Kyoto Protocol and related documents (IPPC, 2001). A measure of the amount of atmospheric CO_2 absorbed by vegetation is known as the net primary production (NPP). NPP is a tool for measuring forest productivity and establishing carbon budgets. The data obtained by calculating NPP can be used as the basis for estimating the impact of both natural disturbances and management activities on forest productivity or assessing the effects of climate change.

Numerous studies have been recently conducted and trials have been performed to produce estimates of forest processes on different spatial and temporal scales using various instruments (i.e. eddy-covariance techniques, satellite images, biogeochemical models, etc.). Among the proposed methodologies, those based on the use of remote sensing data and ecosystem simulation models are particularly promising. In addition, the remote sensing technique and ecosystem simulation models are mostly encouraging. Especially, the remote sensing data have provided the efficient values of vegetation conditions (e.g. LAI, FAPAR, etc.) related to global forest productivity (Waring and Running, 1998 and Maselli et al., 2006). Besides, the ecosystem simulation models are used to effectively combine a variety of data such as the meteorological and soil measurements and the structural and ecophysiological information in order to completely characterize vegetation processes including transpiration, photosynthesis, respirations, allocations and etc.

For the biogeochemical models, the input parameters are combined to use for identifying the physiology, biochemistry, structure and allocation patterns of vegetation functional types or biomes. In case of single-stand simulations, the required model parameters can be measured when spatial coverage increases, data availability decreases, and generalized biome parameterizations are applied. Parameterization may be simplified by using data from the literature in order to make foliar nitrogen across a biome constant. Likewise, allocation of carbon to plant tissues may be assigned as fixed fractions across age classes and climatic zones. Major limitations for regional and global modeling consist of undocumented parameter selection and unknown model sensitivity to parameter variation for larger resolution simulations (White et al. 2000). Admittedly, some ecosystem process models are dynamic and converge towards carbon, nitrogen and water balances. Nevertheless, the models can result in the right answer for the wrong reasons. Besides, a predicted variable such as net primary productivity can have significant error because there are some uncertainties in models.

Due to a lot of variables, it is very important to select suitable inputs for predictive ecological modeling (Faraway and Chatfield, 1998 and Kaastra and Boyd, 1995). Predictive powers of models are varied by specification of inputs. If most of input variables were neglected, the models would loss information significantly. On the other hand, data field collection is both time-consuming and expensive. Generally, the procedure is used to identify the simplest parameterizations or the default variables. Then, it is also applied to test the model in order to determine which parameters need more specification. Another method for parameterization called "data assimilation" begins being applied for scaling ecosystem process recently, even though this method has been used for a long time in atmospheric research (e.g., Cescatti 1997). Data assimilation causes the most consistent model representation with the observation.

1.2 Statement and Rationale of Problem

Carbon fixation is an important process that could balance the global carbon budget. It is commonly evaluated through various observations such as examination of biomes and observation of the Net Primary Production (NPP) (Gamo M and Panuthai S., 2005). NPP measures not only the energy input to the biosphere and terrestrial CO_2 assimilation from the atmosphere, but it can also give information regarding the performance of an ecosystem and the status of a wide range of ecological processes (ORNL DAAC, 2009). Therefore, NPP has oftentimes been remarkably pointed out as a fundamental ecological variable. However, certain limitations of existing field data could impede one's capability of understanding NPP as well as controlling NPP in the forest ecosystems (Clark et al. 2001 and Gower et al. 2001). The reason for this could be the difficulty in directly measuring the above-ground and below-ground biomass increments (Vogt et al, 1996; Silver, 1998). Furthermore, sampling and measuring NPP over a large area could also be difficult over a large area could also be difficult to undertake in spite of the availability of accurate measurements for biomass increments (Wang et al, 2003).

There are several ways of estimating the terrestrial NPP depending on the types of plants and the available measurement tools such as applied satellite data and ecosystem process models, specifically in studying the NPP in a wide area. Such ecosystem process models include the BIOME-BGC or the BioGeochemical Cycles Model, which is one of the well-known models used in forest studies. BIOME-BGC has been applied in finding the NPP of different types of forest ecosystems in various areas of the world. Meng, Wu and Zhou (2005) for example, combined BIOME-BGC, remote sensing and climate model, in estimating the NPP over the entire terrestrial land of China. Nonetheless, one crucial limitation of the BIOME-BGC model is the accuracy of the ecophysiological input parameters for different types of forests because determining the parameter values for modeling in a certain place can be difficult and time consuming (Running 1994). For the deciduous forests, the ecophysiological input parameters are mostly available only for temperate areas such as the Pacific Northwest (Hessl, Milesi, White, Peterson and Keane, 2004). However, to our knowledge, it is rare to find the ecophysiological input parameters of the BIOME-BGC model for tropical deciduous forests.

Tropical deciduous forest plays a significant role in the global carbon budget (Yoshifuji 2006). Whittaker and Likens (1975) reported that tropical forests including deciduous forest could have a mean NPP which is more than the other plant communities in the terrestrial ecosystem. The estimation of NPP in tropical deciduous forest has however, received much less attention than in temperate and boreal deciduous forests (White et al. 1997; Wison and Baldocchi, 2000; Barr et al, 2004). Considering the few but incomplete applications of the BIOME-BGC model for correct parameterization of the ecophysiological input parameters for the accuracy of the modeled NPP in tropical deciduous forests (Aber 1997, White et al. 2000), we chose the Teak (*Tectona Gandis* Linn. F.) for simulating the NPP. An important species in tropical areas, Teak is a fast growing tree species in tropical deciduous forests and widespread in tropical monsoon Asia including southern and central India, Myanmar, Lao PDR and northern Thailand (Bebarta. K.C.,1999).

At the onset, a primary attempt was directed towards the development of a calibration procedure in order to be able to adapt a model for tropical deciduous forests. Applications and reviews of typical inversion algorithms used in ecosystem model calibration found for instance, in Wang et al. (2001, 2006), Knorr and Kattge (2005), Williams et al. (2005), and Raupach et al. (2005), Migliavacca M et al (2009), were used as references and as basis for the development of such procedure. Since the accuracy of modeled NPP is dependent on a correct parameterization of the plant ecophysiological parameters such as C:N ratios, canopy light extinction coefficient, fraction of leaf N in *Rubisco* which had been assumed to be constant for certain species or even plant-functional types (e.g. evergreen needleleaf forest, deciduous broadleaved forest, etc.), several applications described at continental levels (Vetter et.al 2008) were also used as reference. Considering that the aforementioned assumptions do not include parameters from deciduous broadleaved forest in tropical areas such as the Teak tree species, it was proposed in this study to incorporate the measured observations into the ecological process model outputs (e.g. leaf area index, LAI) in the data assimilation, and to optimize the values of one or more unknown model parameters (e.g. allocation ratio).

Data assimilation was incorporated in the study to reduce the impact of incorporating measurement data when such data are unavailable or insufficient, many powerful search and optimization procedures (Migliavacca M et al., 2009 and Srinuandee P et al., 2004) such as the Genetic Algorithm or GA could also be used as an optimization technique. However, GA techniques are different from other optimization techniques in a number of significant ways (Goldberg, 1989). GA consists of three basic operators, namely: selection, crossover, and mutation. The three processes are repeated for many generations to produce the best individual that can represent the optimal solution to the problem. GA has been observed to be robust and appropriate in a wide range of situations to solve optimization problems where ordinary search and optimization techniques are difficult to achieve. It is suitable for solving complex problems in the real world systems. In several studies, GA has been highly applied with crop models such as the coupling SWAP-GA to identify the SWAP unknown parameters (Chemin et al., 2005; Chemin and Honda, 2006; Ines et al., 2006; Akhter et al., 2007; Ines and Mohanty, 2008). However, there are few difficulties and limitations associated with the application of GA with complicated models. It was therefore very challenging to combine the BIOME-BGC with GA and incorporate the performance of genetic algorithm in determining the unknown parameters of BIOME-BGC model using external source of data that have been optimized by the fitness function.

In this dissertation, a data assimilation technique was developed for BIOME-BGC by coupling BIOME-BGC with GA (BIOME-BGC-GA). The coupling was to allow assimilation of remotely-sensed time series (SPOT-LAI) into the ecosystem process model for assimilation of the SPOT-LAI data into the process model. For this purpose, the BIOME-BGC-GA was inverted against the SPOT-LAI in order to retrieve the key drivers of the modeled NPP for tropical deciduous forests. Evaluation of the accuracy of the model was carried out daily and yearly, where the NPP values were compared with those from available literatures, from the MODIS NPP as well as from actual field measurements.

1.3 Research Hypotheses

The following hypotheses are postulated for this study:

(1) The complexities in the field can be determined by using the biogeochemistry model wherein input data are plausible and expect range of values.

(2) The distributed model parameters can be derived from remote sensing observations through the process of data assimilation by exploring dependency of the observed leaf area index to the physical and non-physical properties of the system.

(3) Genetic Algorithms can be applied to implement data assimilation.

1.4 Objectives

The general objective of this research is to develop methodology that could investigate improved carbon fixation in tropical deciduous forest area in Thailand by BioGeochemical Cycles model via the advanced spatial information including remote sensing technique and an optimization technique based on natural genetic called Genetic Algorithm.

The specific objectives are as follows:

1. To develop data assimilation technique for BIOME-BGC using remote sensing data.
2. To find optimized ecophysiological parameters of a deciduous broadleaf forest (Teak forest) in BIOME-BGC model.
3. To simulate and validate the Net Primary Production (NPP)

1.5 Scopes and limitations

The application and validation of the model are limited to one pixel of SPOT – Vegetation (SPOT-VGT) and MODIS NPP. Only one type of biome which is teak or *Tectona grandis* Linn. F. in tropical deciduous forest has been considered in the applications. The case study was conducted in Lampang province, northern of Thailand.

1.6 Study area

The test site is in the Mae Moh Teak (Tectona grandis Linn. F.) Plantation of the Forest Industry Organization (FIO) located in Lampang Province, Thailand at 18°25′ N, 99°43′ E and about 380 m above sea level (Figure 1). During the study period from 2004 to 2007, the annual temperature and precipitation in the study site were 29.8 °C and 1,226 mm, respectively. The dry season in the area could be classified into two: the cool dry season and hot dry season (Yoshifuji, 2006).

The cool dry season (November to February) brings relatively low air temperatures while the hot dry season (March to April) which follows the cool dry season brings high air temperature and vapor pressure deficit (VPD). Tropical monsoons influence the weather in the area by producing great seasonal change in the precipitation. The wet season in the area is from May to October, lasting for six months due to the influence of the south-west monsoon. The mean monthly precipitation during the six-month wet season is 1,120 mm/month. Mean monthly precipitation below 100 mm was however, observed in the dry season from November to April during the 4-year study period.

The study site is 10.64 km^2, which is covered by almost homogenous 40-year old teak forest. The density of the plantation is 360 trees/ha, with mean height of 17.2 m and diameter at breast height (DBH) of 19.5 cm. As observed during the study period, the understory of the forest has been sometimes slashed and often burned in the dry season. The meteorological and some field data for this study were sourced from the AsiaFlux project, which has been conducting flux observations in various places in Asia including in the study site where the project has established a flux tower (Figure 1). The project has been recording and measuring the flux data in the study from 2000 until the present.

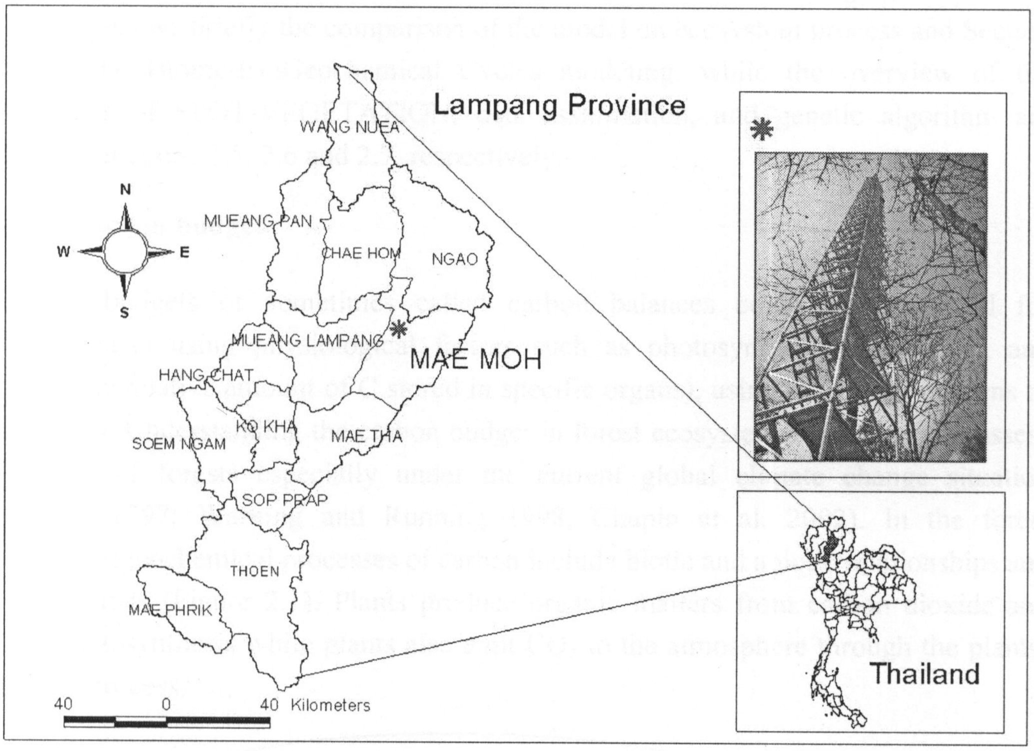


Figure 1.1 Mah Moh study site, Lampang Province, Thailand and position of Flux Tower

CHAPTER II

LITERATURE REVIEW

2.1 General

This chapter begins with a review of the basics of the stated problems and the postulated hypotheses indicated in the previous chapter. It is formulated to illustrate the use of data assimilation for ecophysiological parameterization of BIOME-BGC model in order to assess the net primary production for carbon fixation in the forest biome. This chapter consists of several sections that include the overview of forest carbon budget (Section 2.2). Section 2.3 reviews briefly the comparison of the model on ecosystem process and Section 2.4 shows the Biome-BioGeochemical Cycles modeling, while the overview of the characteristics of SPOT-VEGETATION, data assimilation, and genetic algorithm are presented in Sections 2.5, 2.6 and 2.7, respectively.

2.2 Forest carbon budget

Carbon budgets or sometimes called carbon balances could be calculated for individual plants using physiological factors such as photosynthesis, respiration, and allocation (the relative amount of C stored in specific organs), using time steps in terms of hours or days. Understanding the carbon budget in forest ecosystem is necessary to assess the function of forests especially under the current global climate change situation (Schlesinger 1997; Warning and Running 1998; Chapin et al. 2002). In the forest ecosystem, biogeochemical processes of carbon include biotic and abiotic relationships and their interactions (Figure 2.1). Plants produce organic matters from carbon dioxide and water by photosynthesis while plants also emit CO_2 to the atmosphere through the plants' metabolism process.

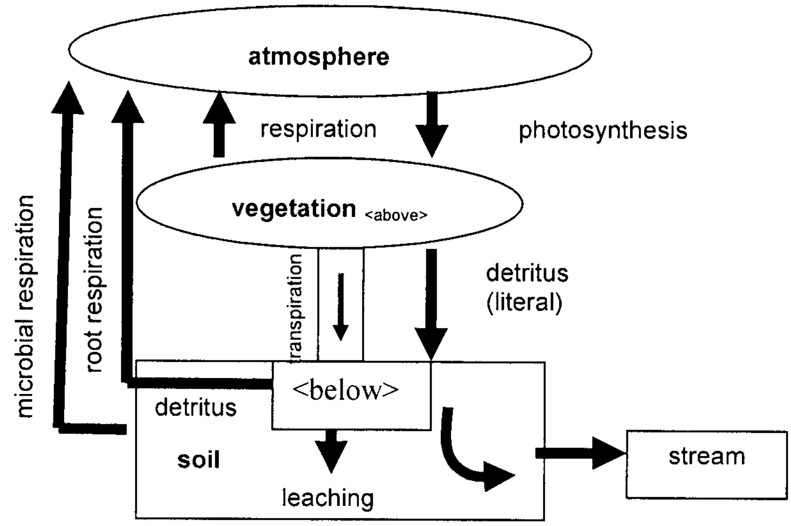


Figure 2.1 General outlines of carbon budget and cycling in forest ecosystem (Shibata et al. 2005)

Photosynthesis in vegetation is the process of injecting carbon into the vegetation and soil ecosystem. The photosynthesis rate of a single leaf is mainly controlled by various environmental factors that include light, air, CO_2 concentration, water, nutrients, and so on. Net photosynthesis is the balance of photosynthesis and respiration from leaf, shoot, branch, trunk, and roots. Plant respiration (autotrophic respiration) includes two major components which could be used to express the productivities in the ecosystem such as Gross Primary Production (GPP) and Net Primary Production (NPP).

Considering the balance of CO_2 as shown in Figure 2.2, on an annual time scale the net exchange of carbon dioxide between a terrestrial ecosystem and the atmosphere (NEP) can be defined as the difference between the GPP and the Ecosystem Respiration (R_{ec}) as indicated in Equation 2.1 (Gamo and Panuthai 2005).

$$\mathrm{NEP} = \mathrm{GPP}-R_{ec} \qquad (2.1)$$

NEP is the net ecosystem production, which is the difference of the CO_2 absorption and emission of a community. If the value of NEP is positive then the ecosystem is considered a sink while if NEP is negative then the ecosystem is a source. Ecosystem respiration R_{ec} is the total respiration of plants (R_a) and non-plant matters (R_h). The value of R_h indicates the respiration by non-plant activity, such as respiration mainly due to the decomposition of soil organic matters. The value of GPP or the gross primary production signifies the amount of photosynthesis. The expression $\mathrm{NEP}=\mathrm{NPP}-R_h$ could therefore be derived from Equation (2.1).

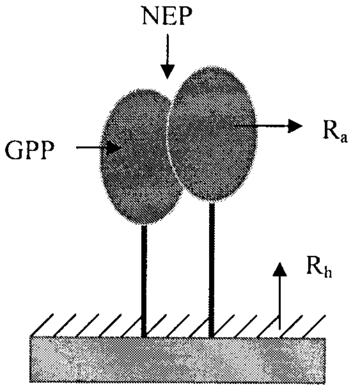


Figure 2.2 carbon flux process in terrestrial ecosystem

$$\mathrm{NEP}=\mathrm{NPP}-R_h. \qquad (2.2)$$

NPP which is the net primary production and expressed as GPP-R_a, could also be described by the following equation:

$$NPP=\Delta B+L+C, \tag{2.3}$$

where ΔB is growing rate of biomass, L, the litter production (litter fall), C the consumption by insects.

2.2.1 Definition of Net Primary Production

Net primary production (NPP) is defined as the net flux of carbon from the atmosphere into green plants, such as the amount of vegetable matter produced. NPP is a tool for measuring forest productivity and establishing carbon budget. The data obtained by calculating NPP can be used as the basis for many issues such as estimating the impact of activities on forest productivity, assessing the effects of climate change on forests and assessing the role that these forests can play in achieving greenhouse-gas reduction.

Net primary production (NPP) is the most important index of plant productivity, related to plant adaptation and crop and woody yields available for consumers:

$$NPP = GPP-R_a \tag{2.4}$$

Table 2.1 shows the global total values of area, biomass and NPP for major biomes. (Saugier et al. 2001)

Biome	Area (10^6 km^2)	Biomass (Pg C)	NPP (Pg C yr^{-1})
Tropical forests	17.5	340	21.9
Temperate forests	10.4	139	8.1
Boreal forests	13.7	57	2.6
Arctic tundra	5.6	2	0.5
Mediterranean shrublands	2.8	17	1.4
Crops	13.5	4	4.1
Tropical savanna and grassland	27.6	79	14.9
Temperate grasslands	15	6	5.6
Deserts	27.7	10	3.5

2.2.2 Production in tropical forest

Whittaker and Likens (1975) reported that the mean net primary production (NPP) of tropical rainforests is more than those of the other plant communities in a terrestrial ecosystem. In Table 2.1, Saugier et al. (2001) showed that the mean NPP in tropical, temperate and boreal forests are 21.9, 8.1 and 2.6 Pg C yr^{-1}, respectively.

2.2.3 Monitoring and analytical method of carbon dynamics

Limitations of the existing field data could however impede the progress in understanding NPP and the control of NPP in forest ecosystems (Clark et al. 2001 and Gower et al. 2001). In spite of the enormous amount of ecological papers on this topic, estimates of forest NPP are sometimes reported based only on incomplete and inappropriate field measurements. Moreover, some hindrances in understanding NPP could also include the substantial efforts for NPP field studies, the challenges of unresolved methods, and a regular lack of conceptual clarity. However, there are several ways of estimating the terrestrial NPP in a wide area depending on the type of plants and available measurements such as through the application of satellite data and models.

2.2.3.1 Field and flux measurement

Observation from the eddy fluxes on forest canopies could be used to quantify the NPP following Equation 2.3. In order to estimate the increment of the stand volume of vegetation in a certain research area, the measurement of the annual increment of the diameter of the breast-height (DBH) and tree height are usually taken. The algometric function of the DBH, tree height and biomass of each component (leaf, branch, root, etc.) of the vegetation are the used to estimate the annual increment (ΔB) of woody tissues of a tree (Hiura 2005).

The litterfall from above-ground vegetation is measured using litter traps set up in certain collecting areas. Good quantification for measuring the litterfall at several points in the study area is essential as these could be considered as replicates, considering that the amount of litterfall largely varies with space.

Observation of the NPP in below-ground vegetation could be difficult to undertake due to some methodological reasons (Smit et al. 2000a). A sequence sample of soil core, in-growth core (Tripathi et al. 2005) and minirhizothron (Majdai et al. 2005; Smit et al. 2000b) are therefore often used to estimate the rate of production and mortalities of fine roots. However, disturbances during *in situ* sampling and measurement of roots sometimes cause serious uncertainties for the quantification of root dynamics (Smit et al. 2000b).

Clark et al. (2001) provided a conceptual framework to guide researchers in their studies toward improving the estimates of forest NPP. For the total NPP field studies, Clark et al. (2001) defined NPP quantity as the sum of the classes of organic materials that should be measured or estimated, and discussed the above- and below-ground components of NPP and the available methods for measuring such components in the field. In their study, the implications of the limitations of past studies for understanding the NPP in forest ecosystems were assessed, and discussed how field NPP measurements can be used to complement tower-based studies of forest carbon flux. In addition, design criteria for future field studies of forest NPP were also recommended.

Clark et.al (2001) also synthesized the data in primary literatures on NPP in old-growth tropical forests and came up with a consistent data set on NPP for such forests. In their studies on the biome, only a few NPP components, all above-ground were however addressed. Given the limited scope of the direct field measurements, they looked for the relationships of the existing data that allow the estimation of unmeasured aspects of production, with those that could be more easily assessed. Their study found a predictive relationship between the annual litterfall and the aboveground biomass increment. For 39 diverse tropical forest sites, they then developed consistent and documented estimates of the upper and lower bounds around the total NPP to serve as benchmarks for calibrating and validating the biogeochemical models with respect to such biome. They developed these estimates based on existing field measurements, current understanding of aboveground consumption and biogenic volatile organic carbon emissions. Based on their findings, we observed that the belowground production is bounded by the range 0.2-1.2 x ANPP (aboveground NPP). Across this broad spectrum of tropical forests (dry to wet, lowland to montane areas, nutrient-rich to nutrient-poor soils), our estimates of the lower and upper bounds on total NPP range were from 1.7 to 11.8 Mg C/ha yr^{-1} (lower bounds) and from 3.1 to 21.7 Mg C/ha yr^{-1} (upper bounds). Clark et.al (2001) also showed that the two relationships that have been used for estimating NPP (the Bray-Gorham relationship based on leaf litterfall and the Miami model based on temperature or precipitation) could not be found in tropical forests.

2.2.3.2 Remote Sensing Observation

Satellite remote sensing of terrestrial vegetation could be used to detect the amount of chloroplast or its activity by processing the "vegetation indexes". Although it could be difficult to estimate NEP directly, some modeling of the NPP could be undertaken using vegetation index.

Goward et. al. (1985) showed that vegetation indices, such as Normalized Difference Vegeta-tion Index (NDVI) are related to net primary production (NPP, g m-2 yearl). Monteith (1987) suggested that NPP under non-stressed conditions is linearly related to the amount of photosynthetically active radiation (PAR, MJ m-2) that is abosrbed by green foliage (APAR, MJ m-2). Further, Kumar and Monteith (1981) showed how the fraction of PAR absorbed (fAPAR) relates to the ratio of red reflectanceto near infrared (NIR). Asrar et.al. (1984) subsequently related the NDVI to the Fapar; hence NDVI may be used to estimate NPP at global scale.

Eck and Dye (1991) described a simple, physically based, satellite remote sensing method for estimating IPAR that uses ultraviolet (UV) reflectivity data from the Nimubus Total Ozone Mapping Sepectrometer (TOMS). Subsequently, Dye and Gward (1993) also created a global APAR image using spectral reflectance measurements from the NOAA-7 AVHRR and TOMS data. Hunt (1994) suggested that global estimates of NPP based on vegetation indices should include a classification among established forest, young forest and non-forest ecosystems to account for difference in zone. To address this problem, Hooda and Dye (1995) developed an automated technique for the identification of agricultural areas using NDVI-climatological modeling.

Rasib et.al. (2007) illustrated the estimation of MODIS NPP using micrometeorological approaches for Peninsular Malaysia and Pasoh Forest Reserve in 2004. They applied continuous field modeling of the tropical rain forest using MODIS satellite data especially for the Pasoh Forest Reserve which has an area of approximately 600 hectares. The model used in such study was successful in estimating the annual above-ground NPP using MODIS satellite data. The approach the applied could be considered as a straightforward scientific method which was significant to estimate the NPP from satellite data. Nonetheless, the results obtained are still being processed taking into account the shortage of recent accurate ground measurements that could be used to validate the findings from their study.

Chhabra and V.K. Dadhwal (2004) report here estimates of monthly net C fixation and net primary productivity over India and its eight regions, using SPOT-VEGETATION 10-day NPP composites, and comparing the monthly patterns of NPP and NDVI. Although many studies use calendar year for reporting NPP, they have adopted an agriculture year (June 1998–May 1999) to better represent the effect of monsoon and the role of agroecosystem in controlling NPP over India. The preprocessing of data included reprojection of the original dataset from geographic coordinates to Albers Equal Area Projection using ENVI 3.5 Image Processing software and computing per pixel and regional NPP.

Furumi et al. (Year) examined the Net Primary Production (NPP) using data from satellites. In order to take the measurement of reflectance (wave-length ranged from 520 nm to 920 nm), they used a helicopter to fly over a cedar forest in Nara, Japan in July 2002. The sensor installed in the helicopter was able to record the Bi-directional Reflectance Distribution Function (BRDF) of the said cedar forest as well as the reflectance of several adjacent areas from the lowest point of the sensor. They also examined the relationship between the reflectance and the sensor observations at the opening and solar illumination angles. When the opening angle is zero, the reflectance is highest because of the effect of less shadow. Taking into account such relationship and using the developed PDM (Pattern Decomposition Method), the VIPD from the reflectance was then calculated and the NPP was then estimated from the VIPD value. The average NPP per month in July 2002 of the cedar forest in Nara, Japan was estimated to be 0.338 kg CO_2/m^2.

2.2.3.3 Model

Biogeochemical models could be one of the main tools for assessing the net exchange of carbon between the terrestrial biosphere and the atmosphere by their ability to fix Equations 1 and 2 at continental and global scales (Melillo et al. 1993; Foley et al. 1996; Bonan 1998; Cramer et al. 1999). These models account for the diversity and complexity of nature by dividing the terrestrial biosphere into broad vegetation classes, such as plant functional groups, as defined by their functions and climate (Holdridge 1967; Woodward 1984; Bonan 2002). The type and amount of vegetation present in a particular location could be evaluated either diagnostically using remote sensing information derived from satellites (Running et al. 1999) or prognostically using dynamic vegetation models (Haxeltine and Prentice 1996; Foley et al. 1996). The results could then be used to parameterize the algorithms in order to determine the photosynthesis and respiration algorithms (Hunt et al. 1996; Bonan 1998; Cramer et al. 1999). Finally, the photosynthesis and respiration algorithms could be evaluated as a function of the environmental variables such as light, temperature and soil moisture. Such factors could be derived either by interpolating the weather measurements or numerically using the weather/climate prediction models.

Recent methods used to evaluate NPP involve the breaking up of productivity into independent parameters such as the incoming solar radiation, radiation absorption efficiency and conversion efficiency of absorbed radiation into organic matter (Kumar and Monteith 1981). The models developed through their study could be considered as advancements over the previous statistical models for the various steps in the productivity build-up process had been properly accounted for in their models.

Generally, the models producing the C budgets are called process models, as they describe the process underlying the system under study. The models are quite useful for investigating certain aspects of C budgets, but they are generally less accurate within observed limits and more expensive for policy analysis than the empirical models based on biomass.

2.3 Model of forest ecosystem process

2.3.1 Overview

Biogeochemical dynamics refer to the interactions among the biological, geological, and chemical components of the Earth's environment. These dynamics are influenced by the interactions between organisms and their surroundings, including soil, sediments, rocks, water, and air, that are the center for understanding the various changes, such as:

- global changes in the cycle of atmospheric greenhouse gases, including carbon dioxide, methane, and nitrogen compounds
- loss of biological diversity from the accelerated cycle of the elements, especially in association with changes in land use
- forest die-back from increased deposition of nutrients and acids from the atmosphere
- degradation of water quality from increased inputs of nutrients through the acidification of streams and lakes from atmospheric deposition

Biogeochemistry models have been used to determine the effects of climate change on forests through the simulation processes of the ecosystem (photosynthesis, transpiration) resulting from tree biomass growth, death and organic matter decomposition with fully-implemented nitrogen, carbon and water cycles. Well-known examples include the PnET (Photosynthesis and Evapotranspiration, Aber et al. 1997) and BIOME-BGC (Thornton et al. 2002), etc. The PnET simulates the monthly output, while BIOME-BGC simulates daily fluxes. Furthermore, PnET could also store carbon, nitrogen and water at specified locations provided appropriate weather data are available, physiography information and ecophysiological traits of the vegetation (Thornton et al. 2002). On the other hand, the BIOME-BGC model predicts the leaf flush and leaf senescence as functions also of available weather data.

2.3.2 Ecological model comparison

Many models are used in studying the ecosystem, and some research studies attempted to establish the differences of such models. Certainly, these models were not designed to determine which models are "best" for diagnosis (i.e., explaining the current function of the models) or prognosis (i.e., predicting the response of the models to future conditions), but rather, to harmonize the similarities and clear up the differences among the models and their components in order that all models could be improved.

Carbon Models Reviewed

- BIOME-BGC: The BIOME-BGC (BioGeochemical Cycles) model simulates the NPP for multiple biomes. Since NPP is computed as the difference between the simulated GPP and autotrophic respiration, the environmental controls operate on the process of photosynthesis as well as respiration. Although nitrogen dynamics have been added, BIOME-BGC relies primarily on the hydrologic cycle and how water availability controls the uptake and storage of carbon (C). The response of NPP to elevated CO_2 is determined mainly by the changes in transpiration associated with reduced leaf conductance, rather than feedbacks from nutrient cycles. BIOME-BGC has a daily time step but without explicit spatial scale. The model has an intermediate number of vegetation and litter/soil pools.

- Century: The CENTURY model simulates carbon, nutrients, and water dynamics for different types of ecosystems. CENTURY includes a soil organic matter/decomposition sub-model, a water budget sub-model, two plant production sub-models (grassland and forest), and functions for scheduling events. The model computes the flow of carbon, nitrogen, and (optionally) phosphorus, and sulfur through the model compartments. Although these four elements have identical organic matter structure, they differ in terms of inorganic compounds. Carbon uptake in CENTURY is controlled primarily by the availability of nitrogen. The grassland/crop and forest production sub-models assume that the monthly maximum plant production is controlled by moisture and temperature, and that the maximum plant production rates depend on the availability of nutrients. Using a monthly time step, the CENTURY model has the finest partitioning of the soil/litter and vegetation pools.
- TEM: The Terrestrial Ecosystem Model (TEM) is a process-based ecosystem model that describes the carbon and nitrogen dynamics of plants and soils for terrestrial ecosystems. This model simulates the limitation of GPP by a multiple of factors, and since plant respiration is explicitly modeled, NPP is simulated as the difference between the GPP and carbon respiration. TEM explicitly simulates the nitrogen mineralization and immobilization dynamics. However, TEM does not consider the influence of vapor pressure deficit on the stomatal conductance or photosynthesis. TEM uses spatially referenced information on climate, elevation, soil, vegetation and water availability as well as soil- and vegetation-specific parameters in order to determine the monthly estimates of carbon and nitrogen fluxes and pool sizes. The response of NPP to elevated CO_2 in TEM is handled by controlling the availability of nitrogen on carbon uptake and storage. TEM operates on a monthly time step and is considered as a global model with a spatial resolution of 0.5 degrees latitude/longitude. This model uses relatively few compartments, with only one carbon pool each for vegetation and soil/litter (two for nitrogen).
- PnET: The PnET models provide a nested set of modular approaches in simulating the carbon, water and nitrogen dynamics of forest ecosystems. The forms of the different versions of PnET are modular and built from the simplest to the most complex versions, however, the algorithms such as for photosynthesis are identical among the model versions. PnET-Day uses foliar mass, specific leaf weight, foliar N concentration, temperature, and radiation flux to predict daily gross and net photosynthesis of the whole forest canopies. In PnET-II carbon allocation and respiration as well as a full water balance are added to predict the NPP, transpiration and runoff. In an empirical soil respiration, PnET-II allows the prediction of carbon balance in the total ecosystem under ambient conditions. This version has been used to predict the combined effects of climate change and

increased atmospheric CO_2. In PnET-CN, compartments for woody biomass and soil organic matter as well as algorithms for biomass turnover, litter and soil decomposition are incorporated to allow for the computation of complete carbon and nitrogen cycles. The original PnET uses a monthly time-step while the PnET-Day uses a daily time-step. Although the PnET models do not have an explicit spatial scale, it could be viewed as regional.

- LoTEC: LoTEC is a mechanistic soil-plant-atmosphere model for carbon storage, and CO_2 and H_2O flux in the ecosystem. Canopy photosynthesis is described by the "Bigleaf" implementation of either the C3 or C4 biochemical model of photosynthesis combined with the sub-model on stomatal conductance. Maintenance respiration for four plant compartments is considered as a function of the tissue nitrogen concentration and temperature, while growth respiration is assumed to be proportional to the change in compartment size. Canopy photosynthesis and maintenance respiration are calculated hourly while carbon allocation, growth, and growth respiration are calculated daily. Litter and soil carbon dynamics are simulated with a monthly time-step. The spatial scale of the model is a half-degree grid cell. This model uses the empirical Miami model, including a factor that represents the response to changing CO_2 as basis for estimating the steady-state of NPP instead of the Farquahar model or other process-based models. Since the Rubisco-limited photosynthesis is not simulated, the use of LoTEC is best justified when light is the limiting factor. A complete run generally requires three phases of simulation, namely: spin-up, historical, and future.
- SiBD: The Simple Biosphere 2 (SiB2) model can simulate the local and regional scale land-surface energy, momentum and mass fluxes using observed forcing ("off-line" mode) or could also serve as the land surface component of the General Circulation Model (GCM). The strength of SiB2 is its vegetation modeling, with dynamic treatment of LAI based on remote-sensed imagery. The meteorology driver data are typically provided at 30-min intervals. Results of the modeling could be provided with a high (in matter of seconds) or low (monthly) temporal resolutions. The spatial scale of each model simulation is the local canopy, but global simulations can also be made by providing separate inputs for each location in a grid. The SiBD model was developed for integration with the GCMs. In SiBD, the dynamic vegetation is simulated using satellite-derived global data from vegetation phenology. Soil hydrological parameterization is then modified to give more-reliable calculations of the inter-layer exchange within the soil profile. SiBD also simulates the gradual changes in surface temperature and reflectance as the amount of snow varies.

This study used BIOME-BGC for the main reason that its general model structure is flexible enough to integrate large scale information, such as regional as well as forest stand level data. This model was selected for our study among other possible alternatives due to its specific suitability in providing information on the water, carbon and nitrogen cycles within the forest and non-forest ecosystems (Running and Hunt 1993; White et al. 2000). BIOME-BGC, however, has established only few and incomplete applications in Southeast Asian areas. Thus, a primary effort was directed to develop a calibration procedure capable of adapting the model to environments, different from those for which it was originally developed.

2.4 The Biome-BioGeochemical Cycles (BIOME-BGC) modeling

2.4.1 BIOME-BGC description

The BIOME-BGC (BioGeochemical Cycles) model is a multi-biome generalization of the FOREST-BGC, a model originally developed to simulate forest stand developments (Running and Gower 1991). The model requires daily climate data and the definition of several key conditions of the climate, vegetation and site to estimate the fluxes of carbon, nitrogen, and water (Figure 2.3) in the ecosystems. In applying the BIOME-BGC, the components of the model should first undergo testing and validation, including the carbon dynamics (McLeod and Running 1988; Korol et al. 1991; Pierce 1993; Running 1994) and the hydrology (Nemani and Running 1989). The model has been successfully applied and validated over a range of diverse biomes, spatial scales and climate regimes including the boreal forests of Alaska and Canada (Keyser et al. 2000; Kimball et al. 1997; 2000; Amthor et al. 2001). The details of the model include applications for multiple biome types and spatial scales (e.g., Thornton et al. 2002; White et al. 2000).

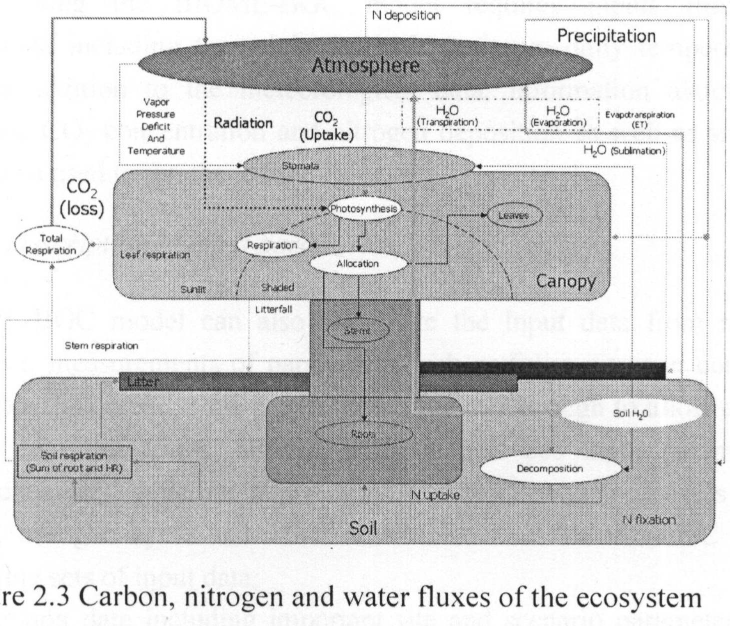
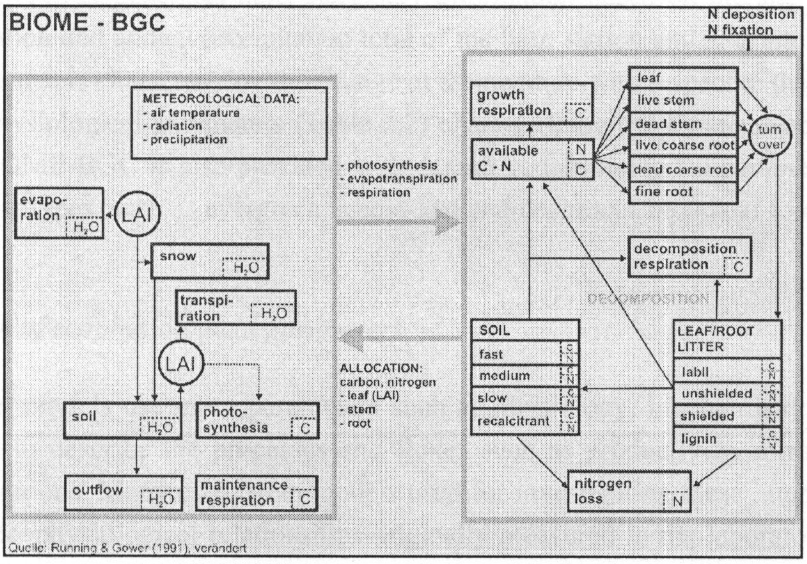


Figure 2.3 Carbon, nitrogen and water fluxes of the ecosystem

The carbon balance portion of BIOME-BGC makes use of daily meteorological data in conjunction with the general stand and soil information to predict net photosynthesis, growth, maintenance and heterotrophic respiration at daily time-step. BIOME-BGC is general in the sense that the surface is represented by singular, homogeneous canopy and soil layers. Detailed descriptions of BIOME-BGC logic are given by Running and Coughlan (1988) and Running and Hunt (1993) as shown in Figure 2.4.



Quelle: Running & Gower (1991), verändert

Figure 2.4 Logic of the BIOME-BGC Model

2.4.2 Model Parameters

Simulation using the BIOME-BGC model requires inputs from a series of meteorological data including the minimum and maximum daily temperatures and daily precipitation. In addition to the meteorological data, information about environmental changes including CO_2 concentration and nitrogen deposition as well as site elevation and soil texture are also used to run the model.

2.4.2.1 Description of input parameters

The BIOME-BGC model can also assimilate the input data from multiple sources (notably plot level measurements of parameters such as foliar nitrogen concentration) and could disaggregate the carbon cycle processes sufficiently enough to allow comparison with a wide variety of observations. BIOME-BGC models use input parameters including physiology, biochemistry, structure, and allocation to describe the processes and fluxes such as productivity, nitrogen cycle and water conditions. The BIOME-BGC model operates using the following sets of input data:

(1) Initialization data including important site and scenario parameters. The key site parameters are elevation, soil texture and effective soil depth. The important scenario parameters include the length of the simulation period, ambient CO_2 concentration (constant or variable), nitrogen deposition, among others.

(2) Existing daily meteorological data series for the simulated site. This data file can be prepared manually or using the MTClim model (Running et al. 1987; Thornton and Running 1999) as inputs in the meteorological data series from a base weather station, which should include at least the daily minimum and maximum temperatures as well as the daily precipitation. MTClim generates other necessary information based on the site parameters (latitude, elevation and annual precipitation total of the base station and site, site aspect and slope) and parameters characterizing the change of temperature with respect to the elevation.

(3) Ecophysiological parameters (Table 2.2) characterizing the biome selected for the simulation. BIOME-BGC is also provided with default ecophysiological parameter sets for the major biome types, such as evergreen needle-leaf and deciduous broadleaf forests (White et al. 2000).

Estimation of ecophysiological parameters

Ecosystem models use input parameters such as physiology, biochemistry, structure, and allocation to describe the processes and fluxes such as productivity, nitrogen cycle, and water conditions. Many ecosystem models used for investigating these interactions are based on the ecophysiological relationships originally measured in the laboratory or field, typically at scales ranging from the leaf to the plot level. These lab- or field-based measurements function for the parameterization and validation of the data sets for the ecosystem models and therefore play crucial roles in current and future model development and implementation.

One of the limitations of the application of the BIOME-BGC model in the field could be the unavailability of accurate physiological input parameters. This is due to the fact that the ecophysiological relationship of forest ecosystems, especially in the Teak species in deciduous tropical forests in Southeast Asia, has not been studied extensively. Moreover, locating such parameter values in existing literatures could also be difficult and time consuming (Running 1994). Although certain data may exist for deciduous forest species, these values are applicable for temperate areas such as the Pacific Northwest (Hessl et al. 2004). Despite this limitation, it is critical that the important parameter values and all the references for such parameter values should be made available for any model-based study (Aber 1997; White et al. 2000).

In biome-based ecosystem models, the commonly measured ecophysiological variables taken from a large number of observations of many communities and locations are typically averaged across the broad vegetation classes (e.g., evergreen needleleaf, broadleaf deciduous, etc.) in order to generate the default parameterization values (Nilson 1995; White et al. 2000). Such default values could include the data collected from low-elevation to subalpine locations, mesic to xeric sites, and recently the disturbed to old-growth forests. Thus, the average or default values could include a high degree of variability even within the broad vegetation types. New parameterization data sets may be required to apply the existing models to new locations, to parameterize new models, or to parameterize existing models more specifically to account for the changes in the physical environment or species.

This study developed a species- and location-specific database of published ecophysiological variables typically used as input parameters for the biogeochemical models of the Teak species or for certain types of deciduous forest ecosystems in Thailand. We selected the parameters based on the requirements of Biome-BGC (White et al. 2000) and on their sensitivity to LAI and NPP outputs, which are described in details in Chapter III. Biome-BGC is a daily time step, spatially independent model that simulates the development of soil as well as plant carbon and nitrogen pools using 43 parameters (Table 2.2). Although the input parameters for this database were investigated based on the structure of Biome-BGC, several other ecosystem models, including Century 5, Daycent, TEM, VEMAP (1995), and PnET (Aber et al. 1996), have been used for determining some of the necessary inputs.

Table 2.2 Ecophysiological input parameters and associated units required to run the Biome-BGC model

Parameter description	**Units**
Woody or nonwoody	Flag
Evergreen or deciduous	Flag
C3 or C4 grass	Flag
Model-defined phenology or user-specified phenology	Flag
Yearday to start new growth (user-specified phenology)	Yrday
Yearday to end new growth (user-specified phenology)	Yrday
Transfer growth period as a fraction of growing season	Proportion
Litterfall as fraction of growing season	Proportion
Annual leaf and fine root turnover fraction	Proportion/yr
Annual live wood turnover fraction	Proportion/yr
Annual whole-plant mortality fraction	Proportion/yr
Annual fire mortality fraction	Proportion/yr
Allocation new fine root C:new leaf C	Ratio
Allocation new stem C:new leaf C	Ratio
Allocation new live wood C:new total wood C	Ratio
Allocation new root C:new stem C	Ratio

Allocation current growth	Proportion
C:N of leaves	kg C/kg N
C:N of leaf litter, after translocation	kg C/kg N
C:N of fine roots	kg C/kg N
C:N of live wood	kg C/kg N
C:N of dead wood	kg C/kg N
Leaf litter labile	DIM
Leaf litter cellulose	DIM
Leaf litter lignin	DIM
Fine root labile	DIM
Fine root cellulose	DIM
Fine root lignin	DIM
Dead wood cellulose	DIM
Dead lignin	DIM
Canopy water interception coefficient	1/LAId
Canopy light extinction coefficient	DIM
All-sided-area to projected-leaf-area ratio	DIM
Canopy average specific leaf area (projected area basis)	m^2/kg C
Shaded SLA:sunlit SLA	DIM
Fraction of leaf N in rubisco	DIM
Maximum stomatal conductance (projected area basis)	m/s
Cuticular conductance (projected area basis)	m/s
Boundary later conductance (projected area basis)	m/s
Leaf water potential:start of conductance reduction	MPa
Leaf water potential:complete conductance reduction	MPa
Vapor pressure deficit:start of conductance reduction	-Pa
Vapor pressure deficit:complete conductance reduction	-Pa

2.4.2.2 Description of the output parameters

The parameter values for the daily and annual algorithms have all been derived, directly or indirectly from the terrestrial ecosystem process model Biome-BGC. The main outputs include the daily and annual (White 2002) data.

Daily

The daily outputs gave the values of GPP (Gross Primary Productivity in kg C m^{-2}), NPP (Net Primary Productivity in (kg C m^{-2}), NEE (Net Ecosystem Exchange in kg C m^{-2}), ET (Evapotranspiration in kg W m^{-2}), OF (soil water outflow in kg W m^{-2}), PRCP (Precipitation in kg W m^{-2}), LAI (Leaf Area Index in m^2 m^{-2}), and LEAFC (Leaf Carbon in kg C m^{-2}).

Annual

The annual data gave the Average Temperature (°C), Precipitation (C m yr^{-1}), GPP (g C m^{-2} yr^{-1}), NPP (g C m^{-2} yr^{-1}), MR (g C m^{-2} yr^{-1}), Max LAI (m^2 m^{-2}), Evapotransporation (cm yr^{-1}), and Outflow (cm yr^{-1}).

2.4.3 BIOME-BGC application

Turner (2006) evaluated the NPP and GPP of MODIS products across multiple biomes. The ground-based NPP and GPP surfaces were generated by applying the BIOME-BGC model. Tatarinov (2006) presented a sensitivity analysis and an adaptation of the BIOME-BGC process model which was thereafter applied to manage forest ecosystems in Central-European conditions. Pietsch (2006) used Biome-BGC to find species-specific parameters for some tree species in central European forests. Kimball (2006) studied the satellite radar remote sensing of seasonal growing seasons for boreal and subalpine evergreen forests. He explained that radar remote sensing measurements of the initiation and length of the growing season corresponded closely with both site measurements. The BIOME-BGC model simulations of these parameters for the sensitivity of the K_u-band scatterometer to snow cover the freeze–thaw dynamics and associated linkages between the initiation of the growing season and the timing of the seasonal snowmelt. Qian et al. (2003) studied the correlation between aircraft fluxes and MODIS NDVI, EVI and LAI. The MODIS GPP, PSN, NPP were validated by the tower measurements at long-term temporal scale and by aircraft measurements at larger spatial scale (Qian et al. 2003).

BIOME-BGC application on deciduous forest

Jolly and Running (2002) utilized an ecosystem process model to investigate the influence of precipitation and soil water potential on the vegetation phenology in semi-arid, drought-deciduous ecosystems in the Kalahari region of South Africa. The timing of the leaf flush was assumed to be the first day during which a rainfall event exceeded that day's estimate of the potential evapotranspiration after a defined dry season. The leaf senescence was assumed to be a dynamic feedback between soil water potential and the net plant carbon gain, and was determined by dynamically modeling the effects of concomitant trends in soil water potential and the net primary production on the leaf area index (LAI). Model predictions of the LAI were compared with the satellite-derived normalized difference vegetation indices (NDVI) for three (3) years at two sites along the Kalahari transect. The mean absolute error for the prediction of the modeled leaf flush data compared with the leaf flush data estimated from NDVI were 10.0 days for the Maun site and 39.3 days for the Tshane site. The correlations between the model-predicted 10-day average LAI and the 10-day composite NDVI for both Maun and Tshane, were high (50.67 and 0.74, respectively, $P<0.001$), suggesting that this method adequately predicts the intra-annual leaf area dynamics in these dry tropical ecosystems.

Turner and et.al (2003) studied the Scaling Gross Primary Production (GPP) over boreal and deciduous forest landscapes in support of the MODIS GPP product validation. The 2001 MODIS GPP product was compared with the scaled GPP estimates (25 km^2) based on ground measurements at two forest sites. The ground-based GPP scaling approach relies on the carbon cycle process model run in a spatially distributed mode. The land cover classification and the maximum annual leaf area index, as derived from Landsat ETM+imagery, were used in order to initialize the model. The model was run using the daily meteorological observations from an eddy covariance flux tower situated at the center of each site. The model-simulated GPPs were confirmed with the daily GPP estimates from the flux tower. The results indicated that at the hardwood forest site, the MODIS GPP phenology started earlier than was indicated in the scaled GPP. As a result, the summertime GPP from MODIS was generally lower than the scaled GPP values. However, the fall-off in production at the end of the growing season was similar to the validated data. At the boreal forest site, the GPP phenologies generally matched because both responded to the strong signal associated with the minimum temperature. Moreover, the midsummer MODIS GPP was generally higher than the ground-based GPP. The differences between the MODIS GPP products and the ground-based GPPs were induced by the differences in the timing of FPAR and the magnitude of light use efficiency as well as by the differences in other inputs to the MODIS GPP algorithm such as the daily incident PAR, minimum temperature, and vapor pressure deficit. It should be noted that ground-based scaling of GPP has the potential to improve the parameterization of light use efficiency in satellite-based GPP monitoring algorithms.

Churkina and Running (2004) used the measurements of extracted timber and modeled forest productivity to investigate the relationship between harvested timber and natural forest productivity under the current conditions as well as the global change scenario. The analysis was confined to coniferous forests and countries that have coniferous forests within their territories. The annual round wood production from the database of the Food and Agriculture Organization (FAO) was used as an approximation of the annual timber harvest for each country. The annual stem primary productivity of coniferous forests was estimated using the BIOME-BGC model. Based on current rates, the annual timber extraction was extrapolated for each country over the next 80 years. In addition, the timber harvest on a country basis was related to the modeled forest stem productivity, assuming that the area of coniferous forests would remain unchanged for the next 80 years. The results of the study suggested that global coniferous forests currently produce more wood than the amount that people can harvest, but this gap is expected to narrow down in the future. However, the results also suggested that wood extraction may surpass forest re-growth by the middle of the next century, even though most coniferous forests are located in high altitudes and having the ability of accelerated stem growth associated with the joint effect of climate change and elevated carbon dioxide concentration in the atmosphere.

Meng et al. (2005) monitored the terrestrial net primary productivity or NPP of China using the BIOME-BGC model based on remote sensing. They used the process-based model BIOME-BGC which is based on the FOREST-BGC model, to simulate the spatial patterns of GPP and NPP over the entire terrestrial land of China. The model was run at ten-day time steps by using 1.0 km Advanced Very High Resolution Radiometer (AVHRR) data of 10-day composite from NOAA satellite series which had been processed on accurate geometric correction, radiance calibration, atmospheric correction and cloud masking, as well as the daily meteorological data from more than 300 weather stations. The differences of the NPP between the various land covers were examined and compared. The results indicated that in 1999, the total NPP of the terrestrial land of China was 1.65×10^9 g C/m^2 with an average of 174.45 g C/m^2. The highest NPP was in the deciduous conifer forests, deciduous broadleaf forests, sparse woods and farmland with average NPP of 346.66 g C/m^2, 318.77 g C/m^2, 309.20 g C/m^2, and 300.47 g C/m^2, respectively, and the lowest was in glaciers, desert and gravel area with average NPP of 12.65 g C/m^2, 14.51 g C/m^2, and 16.73 g C/m^2, respectively.

Tatarinov and Cienciala (2006) presented the results of a sensitivity analysis and an adaptation of the BIOME-BGC process model which was thereafter applied to manage the forest ecosystems in Central-European conditions. Their study described a more specific model adaption for classically managed forest ecosystems, mimicking the thinning and felling regimes in even-age forests of four major tree species (beech, oak, pine and spruce), from which a detailed sensitivity analysis was conducted. They specifically analyzed the effects of site and ecophysiological parameters on the modeled state variables (carbon pools in biomass, litter and soil, and net primary production (NPP)). Results of their analysis revealed a high sensitivity of all tested variables to the following site parameters: total precipitation, rooting depth, sand fraction (for sandy soils only), ambient CO_2, and nitrogen input parameters. Similarly, the tested variables were highly sensitive to the following ecophysiological parameters: leaf and fine root C:N ratio, new stem C to new leaf C ratio, new fine root C to new leaf C ratio, specific leaf area, maximum stomatal conductance, fire mortality and fraction of N in Rubisco (specifically for deciduous species). Moreover, the results also indicated that the whole plant mortality had high effect on carbon pools but with minimal effect on NPP.

2.5 Characteristics of SPOT VEGETATION (SPOT-VGT)

The SPOT (System Probatoire d'Observation de la Terra) series is a land-observing satellite which was launched by the Center National d'Etudes Spatiales. The satellite orbits the earth on the sun-synchronous sub-recurrent orbit at an altitude of 832 km in 10.3 minutes. It has an orbit inclination of 98.7 degrees and a recurrent frequency of 26 days.

Since the first launching of the satellite in 1986, five satellites have been launched so far. The VEGETATION sensor with a spatial resolution of 1.0 km and a swath of 2000 km for the extensive vegetation observation was added to the fourth satellite as well as to the fifth.

Table 2.3 Specifications of the VEGETATION Sensor

Band	Wavelength (μm)	Type	Spatial resolution	Swath
1	0.430-0.470	Blue	1.15 km	2,250 km
2	0.610-0.680	Red		
3	0.790-0.890	NIR		
4	1.580-1.750	SWIR		

The same spot is observed once to several times a day in order to observe two types of data, namely: primary product (P) which provides the raw data and synthesis product (S) which is capable of synthetically processing after corrections are applied such as atmospheric correction. The S10 product was produced by synthesizing the data for ten days. Therefore, the observed data for one year comprised 36 units of information for the S10 data.

The VGT-S10 (ten-day synthesis) products are composite (maximum-value) products. All the segments during this period are compared in order to select the 'best' ground reflectance values. These products provide the data from all spectral bands, the NDVI, and the auxiliary image acquisition parameter data. The continental S10-composite data products (spectral band data, data quality layer, and NDVI) are downloaded through the GIMMS group from Vito, Belgium. The individual composite NDVI data for each period are extracted from the S10-HDF file and post-processed. The post-processing steps include a re-projection from the native global Mercator projection to other projections, continental and regional sub-setting, and the incorporation of flags for bad data, clouds, and land mask.

2.6 Data Assimilation

ECMWF (2002) defined data assimilation as a technique used in analyzing the observed information which had been accumulated into the model by taking advantage of the consistent constraints with the laws of time evolution and physical properties.

Data assimilation is a combination of information from observations and models of particular physical system in order to get the best possible estimate of the state of such system (Swinbank et al. 2004). The technique has wide applications across a range of earth sciences, with its major application being in operational weather forecasts. Others include oceanography, atmospheric chemistry, climate studies, and hydrology. Data assimilation for the Earth system is a comprehensive survey of both the theory of data assimilation and its application in a range of earth system sciences. Data assimilation is a key technique in analyzing the remote sensing observations and thus, is particularly useful for those intending to examine the wealth of measurements from recent research satellites.

2.6.1 The need for optimization

Optimization is the process of adjusting the inputs or characteristics of a device, mathematical process or experiment to find the minimum or maximum output or result. While the mathematical approach of root finding requires searching for the zeros of a function, optimization requires finding the zeros of its derivatives. By definition, finding the root gives the global minimum of a function which is not the case in optimization. Some difficulties could be encountered in determining the root if a given minimum is the best (global) minimum or a sub-optical (local) minimum. Usually, finding the root of some non-linear functions is especially difficult, so that linearization of the various sub-components of the problem is sometimes used. However, in this process the complexity of the problem could be discarded and thus could lead to some collateral impacts on the accuracy of the solution (Yann 2006). The flowchart of the optimization is shown below:

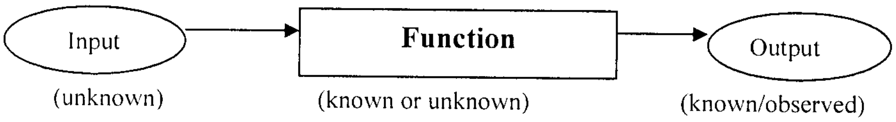


Figure 2.5. Basic problem requiring optimization (Yann 2006)

2.6.2 Ecological model and data assimilation

There are inherent limitations in the measurement and modeling of the ecosystem carbon dynamics. Some measurements are usually patchy in space and discontinuous in time, and modeling of carbon dynamics is always built on a set of principles coupled with assumptions and imperfectly defined parameters. Advanced data assimilation techniques based in statistics or optimization theory can mitigate such limitations through the combination of a series of measurements with dynamic models.

Data assimilation techniques have been proven as vital tools to develop quantitatively realistic ecosystem models, by improving their accuracy, efficiency and prognostic ability. These techniques fix the unknown parameters in the governing equations and save time in finding the parameter sets that can better simulate the data. The models' structures can also be refined based on the correlation between the parameters by using the error covariance matrices (Tziperman and Thacker 1989; Matear 1995). Some models with data assimilation routines developed or still in the process of development include SipNET, TREES,ORCHIDEE, BETHY, TRIFFID, ED, Biome-BGC, LoTEC, SiB3.

Lawson et al. (1995) applied the data assimilation technique to a predator-prey model, based on the so-called adjoint method but allowing the computer code for the adjoint to be constructed directly from the model computer code. This technique is straightforward and reduces the chance of introducing errors in the construction of the adjoint code. The implementation of the technique is demonstrated by applying it to a simple predator-prey model in a model-fitting mode. A series of identical twin numerical experiments were used to show that this data assimilation approach can successfully recover the model parameters as well as the initial conditions. However, easy recovery of the required values is dependent on the form of the model equations as well as on the type and amount of the available data. Additional numerical experiments showed that sufficient coefficient and parameter recoveries could also be possible even when the assimilated data contain significant random noise. Thus, for biological systems that can be described by ecosystem models, the adjoint method represents a powerful approach for estimating the values for little-known biological parameters such as initial conditions, growth rates, and mortality rates.

Matear (1995) developed a simulated annealing optimization algorithm to optimize the parameters of ecosystem models. The optimization was used to directly determine the model parameters required to reproduce the observed data. The optimization routine was formulated in a general manner and modified easily to include additional information on both the desired model output and the model parameters. From the optimization routine, the error analysis of the optimal parameters was produced from the error-covariance matrix which gave both the sensitivity of the model to each model parameter and the correlation coefficients between all pairs of model parameters. In addition, the optimization analysis provided a means of assessing the necessary model complexity required to model the available data. He demonstrated the technique where the optimal parameters of three different ecosystem model configurations were determined from nitrate, phytoplankton, meso-zooplankton and net phytoplankton productivity measurements at Station P. The error analysis of the optimal parameters at Station P indicated that the data are able to resolve up to 10 independent model parameters. This is always less than the number of unknown model parameters, indicating that the optimal solutions are not unique. Thus, the simple nitrate-phosphate-zooplankton ecosystem model was successful in reproducing the observations. This study also justified the use of a more complicated model at Station P but required additional data to constrain the optimization routine. Although there was evidence supporting the importance of the microbial loop at Station P without additional ammonium and bacteria measurements, a more complicated model that includes such processes could not be validated.

Spitz et al. (1998) focused on the feasibility assessment of a data assimilation technique with sparse time series observations such as from the Bermuda Atlantic Time-Series Study (BATS), to estimate the poorly known parameters for the annual cycle of a nitrogen budget model in the upper ocean mixed-layer. They carried out two groups of data assimilation experiments: first, the twin experiments using model-generated observations were run to determine if the frequency of the data collected from BATS was sufficient to estimate all the model parameters; and secondly, a data assimilation of the BATS data from 1988 to 1993 was attempted. Specifically, the pelagic ecosystem model for this study was based on Fasham et al. (1990) model that was previously tried in their study.

This experiment further showed that some of the model parameters could not be independently estimated. This conclusion leads to a simplification of the model and a redefinition of its parameters. Based on the success of the twin experiment in estimating the model parameters, an attempt was made to assimilate actual observations from BATS. Assimilation of the real data led to the conclusion that even though the frequency and types of observations are adequate to estimate the model parameters, the considered model is not appropriate for the annual cycle of the BATS ecosystem.

Vallino (2000) examined the use of data assimilation and mesocosm experiments to facilitate the development of food web models. The components of the data assimilation include the construction of measurement models, the adjoint technique to obtain gradient information on the objective function, the use of parameter constraints, incorporation of discrete measurements, and assessment of the observability of the parameters. He also examined the effectiveness of classic and contemporary optimization routines used in data assimilation. The standard compartment-type food web model was employed with emphasis on organic matter production and consumption. The mesocosm experiments which were designed to examine the interaction of inorganic nitrogen with organic matter provided the data used to constrain the model. Although he was not able to obtain the reasonable fit between the mesocosm data and the food web model, he observed that the model lacked robustness to be applicable in trophic gradients, such as those occurring in coastal environments. The robustness problem could be due to the inherent structural problems that render the model extremely sensitive to the parameter values. Furthermore, parameters governing actual ecosystems were not constant, but rather varied as a function of environmental conditions and species abundance, which increased the sensitivity problem. He concluded by briefly discussing possible improvements of the food web models and the need for rigorous comparison between models and data (a modeling workbench) so that the performance of the competing models could be assessed. Such a workbench could facilitate the systematic improvements in the prognostic marine food web models.

Kuroda and Kishi (2004) applied a data assimilation technique to determine the biological parameters in the PICES (North Pacific Marine Science Organization) prototype lower trophic level model (NEMURO). The North Pacific Ecosystem Model for Understanding Regional Oceanography (NEMURO) has about 80 biological parameters and 11 initial values. They used a sensitivity analysis to choose eight parameters which had the most impact on the relevant simulated values. These parameters were selected as control variables for the data assimilation. The model output, using the optimum parameter values determined by the assimilation, conform to the data better than those obtained from the first guessed parameter values.

They also observed that ecosystem models with more compartments would require finding out more biological parameters since almost all of the values depend on measurements and laboratory experiments. In analyzing or improving the model, they had trouble deciding on which parameters to focus on. Their method was effective because it is unbiased and it was not necessary to change all the parameters. This assimilation used only the control variables that were expected to strongly affect the model output, and still it yielded more than just the optimum values of the control variables. Their results also revealed what parameter values were misestimated which should be changed.

Cossarini et al. (2004) applied the data assimilation techniques to an ecosystem model of the Venice lagoon to obtain reliable and coherent estimations of the chlorophyll and nutrient fields, improving the knowledge on spatial and temporal evolution of the trophic state. They compared the results obtained by simple techniques such as the Direct Insertion and Optimal Interpolation, and more complex, state of art technique such as the application of the Error Subspace Statistical Estimation scheme on the Extend Kalman Filter.

Chen et al. (2008) improved the state-parameter analysis of ecosystem models using data assimilation. They introduced the smoothed ensemble Kalman filter (SEnKF) to estimate simultaneously the state variables and parameters of a forest carbon flux partition model. The SEnKF method substantially and significantly improved the flux estimates of a flux partition model and dramatically reduced uncertainties that stemmed from parameters and driving forces. Simultaneous parameter estimation can use near real-time observations to improve the predictive ability of dynamic models.

2.7 Genetic Algorithm

The theoretical considerations on Genetic Algorithms, parameter estimation by inverse modeling and system integration for regional analysis are discussed in this section. Genetic Algorithm (GA) is an artificial genetic system based on the processes of natural selection and natural genetics. GA involves three operators, namely: reproduction, crossover, mutation. A simplified GA cycle is shown in Fig. 2.6

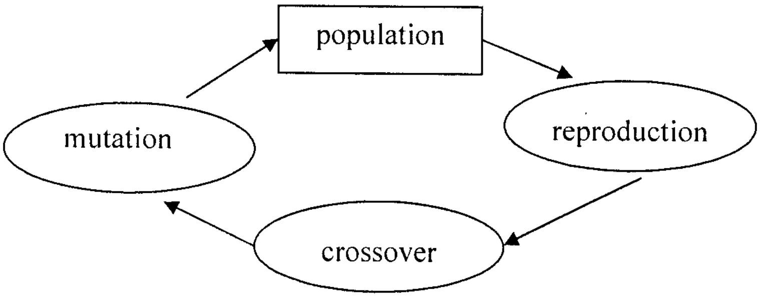


Figure 2.6 GA cycle

2.7.1 Overview of Genetic Algorithms

Genetic algorithms are general purpose-search algorithms inspired by Charles Darwin's principle of 'survival of the fittest' to solve complex optimization problems (Holland 1975; Goldberg 1989). A population of competing solutions evolves over time to converge into an optimal solution, which could be represented by a chromosome consisting of several genes. Genetic algorithms are search algorithms based on the mechanics of natural selection and natural genetics by combining 'survival of the fittest' among a string of structures with a structured yet randomized information exchange to form a search algorithm with some of the innovative flair of human search. Genetic Algorithms (GAs) are adaptive heuristic search algorithm premised on the evolutionary ideas of natural selection and genetic. The basic concept of GAs was designed to simulate the processes in the natural system necessary for evolution, specifically those that follow the principles first laid down by Charles Darwin on 'survival of the fittest'. As such, GAs represent the intelligent exploitations of a random search within a defined search space to solve a problem.

There are numerous variations of Gas but the GA presented here is simple as outlined by Goldberg (1989). As an example, GA was applied to find an optimal set of input variables for the prediction of the presence or absence of benthic macro-invertebrate taxa in un-navigable Flemish watercourses. The chromosomes consisted of 17 genes, each representing an input variable provide with a binary coding. This means that a particular variable was either selected (represented by '1') or not (represented by '0'). Each chromosome of a particular generation is allocated a piece of the roulette wheel, according to their fitness for the prediction of benthic macro-invertebrates. By spinning the roulette wheel, a chromosome is selected for reproduction. In this manner, chromosomes with high fitness have higher chances of being selected for the next generation. This implies that an input variable subset which results in a high performance has a higher probability to be selected. Moreover, crossover is set with a probability of 60% while mutation occurs at a probability of 3%. This low setting of mutation could avoid getting trapped in local optima during the search. The initial population consisted of 20 chromosomes that were evolved over minimal 40 generations were the parameters set after the preliminary experiments.

GAs are different from other more normal optimization and search procedures in four ways where the GAs work with code (Goldberg 1989): (1) GA works by coding the parameter set (string) but not with the parameters themselves; (2) GA searches from a population of points not from a single point; (3) GA uses objective function information not derivatives or other auxiliary knowledge; and (4) GA uses probabilistic transition rules not deterministic rules.

First pioneered by John Holland in the 60s, GAs have been widely studied, experimented and applied in many fields especially in the engineering world. Not only does GAs provide an alternative method to solving problems, it consistently outperforms other traditional methods in most of the problem links. Many of the real world problems involved finding optimal parameters, which might prove difficult for traditional methods but ideal for GAs. However, because of their outstanding performance in optimization, GAs had been wrongly regarded as function optimizer. In fact, there are many ways to view GAs. Perhaps most users resort to the GAs looking for problem solvers but this could be just a restrictive view (De Jong 1993).

GAs were introduced as computational analogy of the adaptive systems, and modeled loosely on the principles of evolution via natural selection, employing a population of individuals that undergo selection in the presence of variation-inducing operators such as mutation and recombination (crossover). A fitness function is used to evaluate the individuals, but reproductive success varies with fitness. The genetic programming works like a flow as shown in Figure 2.7 (Natthaphob 2006).

The most common type of genetic algorithm works like in the evolution of man where a population is created with a group of individuals created randomly. The individuals in the population are then evaluated, where the evaluation function is provided by a programmer who gives the individuals with a score based on how well they perform their given tasks. Two individuals are then selected based on their fitness, the higher the fitness, the higher is the chance of being selected. These individuals then "reproduce" to create one or more offspring, after which the offspring are mutated randomly. This continues until a suitable solution is found or a certain number of generations are produced depending on the needs of the programmer. Based on natural selection after an initial population is randomly generated, the algorithm evolves through three operators, namely:

- selection which equates to survival of the fittest;
- crossover which represents mating between individuals; and
- mutation which introduces random modifications.

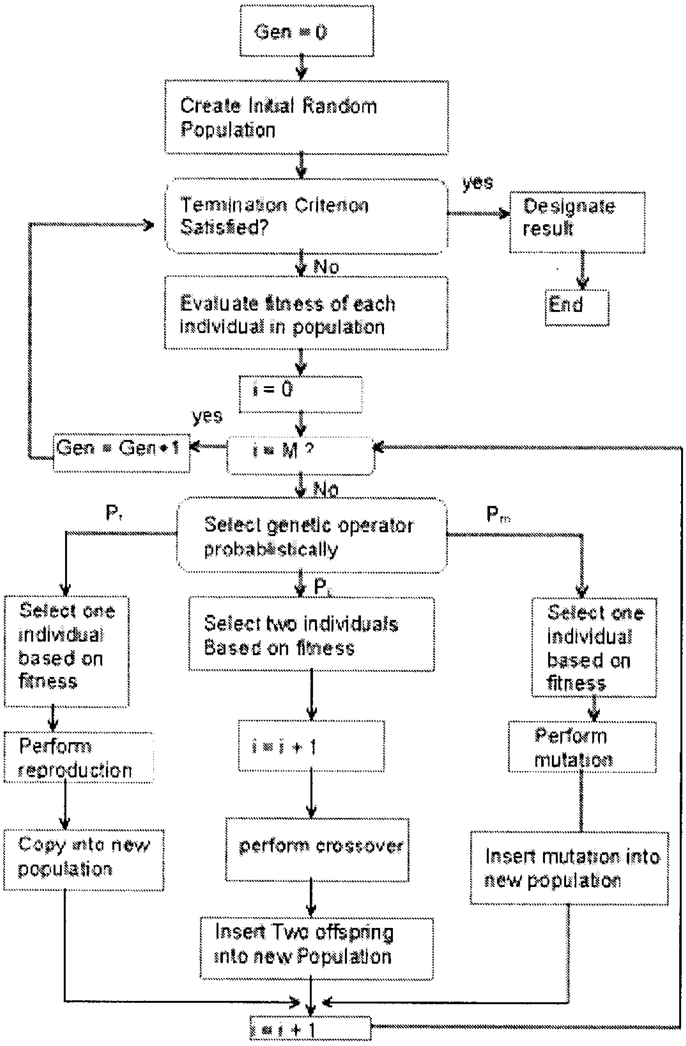


Figure 2.7 Genetic programming tasks (Natthaphob 2006)

2.7.1.1 Selection Approaches

Agents are selected that will make it to the crossover phase. While those that are not selected will die and therefore, their genes will not be passed on to the agents in the next generation. The selection process is based on the probability that agents evaluated with higher Y values will most likely be selected for the next phase. Those with low values will probably not. The key point is that this phase has an element of randomness just like the survival of organisms in nature.

The probability for selection is based on the agent's Y value relative to the rest of the population (survival of the fittest). Selection begins by determining an agent's relative fitness by dividing its Y value by the sum of all the Y values of the agents in the population. Then a random number generator is used to select the agents for the crossover phase. The chance of an agent being chosen during each spin of the random number generator is equal to the agent's relative fitness. The number of agents selected is equal to the population size, it is therefore necessary to keep the size constant for every generation. Some agents will be selected more than once in which case multiple copies of such agent would be present in the set used in the crossover operation.

Many methods are used to select the best chromosomes, such as the roulette wheel selection, Boltzman selection, tournament selection, rank selection, steady state selection, among others. Some of them are described in details as follows:

Roulette wheel selection: In roulette wheel selection, individuals are given the probability of being selected directly proportionate to their fitness. Two individuals are then chosen randomly based on such probabilities and allowed to produce offspring. Imagine a roulette wheel where all chromosomes in the population are placed and everyone has its place accordingly based on its fitness function, as shown in the picture below.

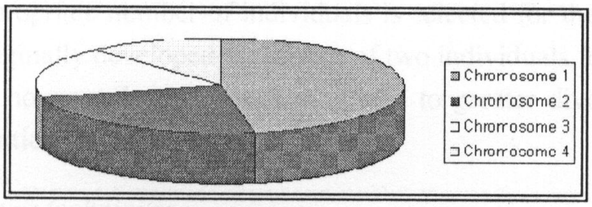


Figure 2.8 Roulette wheel selection process

A marble is then thrown into the pie and the chromosome is selected. Chromosome with bigger fitness will have the chance of being selected a number of times.

Ranking Schemes: The previous selection will have problems when the difference of the fitness is big. For example, if the best chromosome fitness is 90% of the roulette wheel then the other chromosomes will have very few chances of being selected. In rank selection, the population is ranked first and then every chromosome receives fitness from such ranking. The worst will have fitness 1, second worst 2, etc. and the best will have fitness N (number of chromosomes in the population).

The following picture shows how the situation varies after changing the fitness to order number, after which all the chromosomes have the chance of being selected. But this method can lead to slower convergence, because the best chromosomes would not be much different from the others.

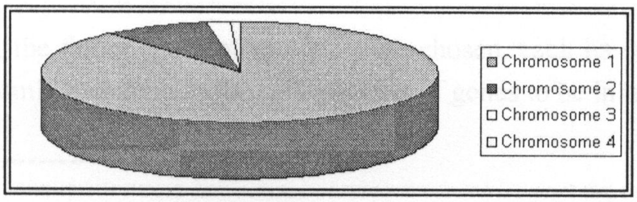


Figure 2.9 Situation before ranking (graph of fitness)

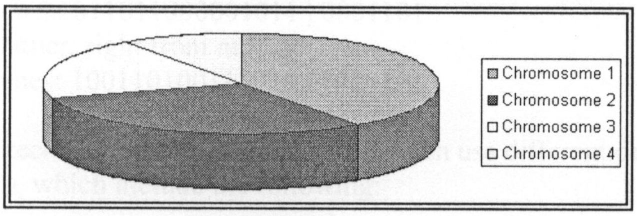


Figure 2.10 Situation after ranking (graph of order numbers)

Tournament selections: Goldberg and Deb (1990) compared the various selection schemes and indicated a preference for the tournament selection. In tournament selection, a group of individuals are chosen at random from the population and the individual with the highest fitness is selected for inclusion in the reproduction process. The procedure is repeated until the appropriate number of individuals is selected for the new generation. The approach has been originally developed for groups of two individuals for the so-called binary tournament selection however larger groups may lead to greater diversity and a smoother progression to the solution.

2.7.1.2 Crossover techniques

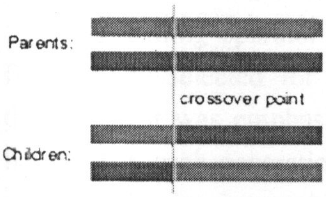
Crossover is the process of combining the genes of one agent with those of another to create offspring that inherit the traits of both parents (Goldberg 1989). A crossover rate is the odds of an agent being selected for the crossover operation. The agents that are not selected will not have their genes changed before proceeding to the mutation phase. Those that are chosen will be paired with a mate, which is another agent which has also been selected for the crossover. From each pair, two offspring will be created that will replace their parents. To determine which genes are inherited from the father and which genes will come from the mother, a random number between one and the total number of genes minus one will be created. For the first offspring, the genes numbered between one and the random number will be inherited from the father. The genes numbered between the random number plus one and the maximum number of genes will come from the mother. The genes for the second offspring will be inherited just like those of the first offspring except that the genes that came from the father in the first offspring will come from the mother and those inherited from the mother will come from the father.

For example, the father and mother agents are chosen, each having 22 genes. Number 15 is created randomly and denotes the split position of genes to be inherited (AI Lab 2006).

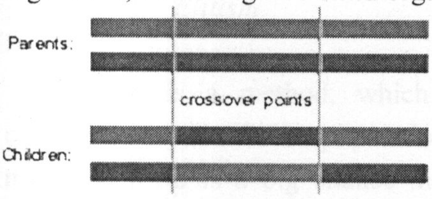
```
Split position ----------------------+
                                     |
Father's genes:     011011000001011 | 1010111
Mother's genes:      100110100111010 | 0001101
                                     |
Offspring A's genes: 011011000001011 | 0001101
- left side from father, right from mother
Offspring B's genes: 100110100111010 | 1010111
```

Many crossover techniques exist for organisms which use different data structures (Eshelman 1991), which include the following:

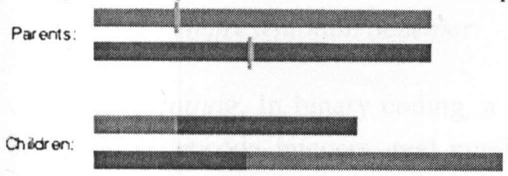
One point crossover: A crossover point on the parent organism string is selected. All data beyond that point in the organism string is swapped between the two parent organisms. The resulting organisms are the children:



Two point crossover: Two point crossover calls for two points to be selected on the parent organism strings. Everything between the two points is swapped between the parent organisms, rendering two child organisms:



Cut and splice: Another crossover variant, the "cut and splice" approach, results in a change in length of the children strings. The reason for this difference is that each parent string has a separate choice of crossover point.



An important aspect of crossover (in binary coding) to multivariate problems is that it occurs only at gene boundaries because each gene consists of alleles or bits. Crossover may split the genes, which may cause either a gain or loss to a GA. This is not the case for real coded GA where the gene comprises a single allele and is itself the parameter value (Whitely 1989; Goldberg 1989; Michalewicz 1996)

2.7.1.3 Mutation

After the selection and crossover, there is a new population full of individuals. Some are directly copied while the others have been produced by crossover. In order to ensure that the individuals are not all exactly the same, only a small chance of mutation is allowed. The loop through all the alleles of all the individuals and if that allele is selected for mutation, either it is changed by a small amount or replacing it with a new value.

Just as in nature, some agents will have random mutations occurring in their genes. The mutation rate specifies the chance that a given gene in an agent will be mutated. If a gene is selected for mutation then its value will be changed. In the case of bit representation, the gene will simply be flipped, that is, a one is changed to a zero or a zero is changed to a one.

2.7.1.4 Niching

The niching operator assures certain diversity in a population, allowing not only the fittest to be selected for reproduction but also the less-fit individuals. In the previous discussion, it was emphasized that population diversity and selective pressure have to be balanced in each generation in order not to attain premature convergence with suboptimal solution. Niching is an advanced GA operator that promotes population diversity through a sharing scheme.

2.7.1.5 Elitism

Elitism is a method, which copies first the best chromosome (or a few best chromosomes) to a new population. When creating a new population by crossover and mutation, there is a big chance that the best chromosome is lost. The rest is done in a classical way. Elitism can very rapidly increase the performance of GA, because it prevents losing the best found solution.

2.7.1.6 Representation Schemes

Binary coding. In binary coding, a chromosome is represented by a string of binary bits that can encode integers, real numbers, or anything else appropriate to a problem. Binary strings are easy to operate and within any gene, binary representations can be mapped to values in a range feasible for the variables represented. Following the operation of reproduction (selection, crossover and mutation), the fitness of a particular chromosome is evaluated after the binary values are decoded back into their original form. However, the standard binary coding of variables permits large gaps in the variable values between generations, which can lead to difficulty in arriving at a good solution (Goldberg 1989; Wardlaw and Sharif 1999).

Read coding (Floating point). Real coded chromosomes have been used with success. In a real coded GA, individual genes of a chromosome are initially the values allocated randomly within the feasible limits of the variable represented. With a sufficiently large population of chromosomes, adequate representation would be achieved. There is a significant advantage in not wasting computer time on decoding the objective function of an evaluation, although a more careful approach to mutation is required. Nevertheless, in real coding there is no discrete decision of the variable space (Haupt and Haupt 1998; Wardlaw and Sharif 1999). Table 2.4 shows the distinction of the representative coding from the above discussion (after Goldberg 1989).

Table 2.4 Comparison of binary and real-coded genes

Real coded	Binary coded
0.0	0000
1.0	0001
2.0	0010
3.0	0011
4.0	0100
5.0	0101
6.0	0110
7.0	0111
8.0	1000
9.0	1001
10.0	1010
11.0	1011
12.0	1100
13.0	1101
14.0	1110
15.0	1111

Genetic algorithms are very effective way of quickly finding a reasonable solution to a complex problem. Granting that GAs are not instantaneous or even close, they could perform an excellent job of searching through a large and complex search space. Genetic algorithms are most effective in a search space for which little is known, for they produce solutions that solve certain problems in ways that may never have even been considered. They can also produce solutions that only work within the test environment and could flounder, once these are used in the real world. Put simply: "genetic algorithms could be used for everything that could not be easily done with another algorithm."

2.7.2 Parameters of GA

This section discussed some basic recommendations for deciding to implement genetic algorithms. Considering that these recommendations are very general, most study would probably want to experiment with their own GA for specific problem, because there is still no general theory which could describe the parameters of GA for any problem.

The recommendations are results of some empirical studies of GAs, which have been often performed only on binary encoding.

- *Crossover rate*

  Crossover rate generally should be high, about 80%-95%. (However some results have shown that for some problems the crossover rate of about 60% could be the best.)
- *Mutation rate*

  On the other hand, mutation rate should be very low. The best rates reported are about 0.5%-1%.
- *Population size*

  It is surprising that very big population size usually does not improve the performance of GA (in terms of speed in finding solution). Good population size is about 20-30, however sometimes sizes between 50 and 100 are often reported as the best. Some research results also showed that the best population size depends on the encoding or on the size of the encoded string. This means that, if a chromosome has 32 bits the population should be 32, but this is surely two times more than the best population size for chromosome with 16 bits.
- *Selection*

  The basic roulette wheel selection can be used, but sometimes the rank selection could be better. There are also some more sophisticated methods which could change the parameters of selection during a GA run. Basically, they behave like simulated annealing, but surely elitism should be used.
- *Encoding*

  Encoding depends on the problem and also on the size for instance of the problem.
- *Crossover and mutation type*

  Operations largely depend on the encoding and on the problem.

2.7.3 Applications of GA

Genetic algorithms have been used for difficult problems (such as NP-hard problems), for machine learning and also for evolving simple programs. They have also been used in art specifically for evolving pictures and music.

The advantage of GAs is in their parallelism. GA is travelling in a search space with more individuals (and with genotype rather than phenotype) so they are less likely to get stuck in local extremes like some other methods.

GAs would be also easy to implement because once a certain GA is available, a new chromosome (just one object) could be written to solve another problem. Using the same encoding, all what is needed is to change the fitness function. However, choosing the encoding and fitness functions could also be difficult.

A disadvantage of GAs is in their computational time as they can be slower than some other methods. But with today's computer capabilities, this is no longer a big problem.

To get an idea about the problems solved by GA, the following is a short list of some GA applications:

- Nonlinear dynamical systems - predicting, data analysis
- Designing neural networks, both architecture and weights
- Robot trajectory
- Evolving LISP programs (genetic programming)
- Strategy planning
- Finding shape of protein molecules
- TSP and sequence scheduling
- Functions for creating images

2.8 Conclusion

Chapter II reviews the fundamentals of the problems and hypothesis stated in Chapter I. The challenge of deriving the accurate input parameters for the effective application of simulation models in the field could be overcome by using data assimilation technique. Data assimilation is a key technique in the analysis of remote sensing observations. GA is an advance data assimilation method that could be applied in this study. The details on the development of the methodologies will be discussed in Chapter III.

CHAPTER III

RESEARCH METHODOLOGY

This chapter described over all methodology and materials that been used to completed the study. It consists of four sections; over all methods, materials, the optimization algorithm base on a genetic algorithm and simulation model.

3.1 Overall Methodology

Pre-processing of remotely sensed data

SPOT S10 scenes of 2004-2007 were georeferenced by identified upper-lower of left and right geographic of image from identification information. The SPOT s10 data were pre-processed to remove all residua cloud contaminations, consisted of a Min-Max filtering applied to remove the atmospheric noise fully described in Sawada, 2002. After geometric correction, The SPOT S10 original data was converted to NDVI and LAI respectively as explained in section 3.2

Calibration phase

An indicator of a model's complexity may be the number of parameters that are used in the model to characterize various ecosystem processes or to represent different environmental properties. Determining the appropriate values for these parameters requires great diligence: White et al. (2000) presented a 40-page referenced source data to calculate a default set of ecophysiological parameters for Biome-BGC (supplied with the distribution of the BGC model). However, such default parameters are intended for general guidance only. For a model as complex as Biome-BGC, small uncertainties in the parameters may lead to the generation of a wide range of variability in the subsequent simulations. For particular applications, therefore, model parameters should be calibrated against site-specific measurements to ensure the quality and accuracy of the results of the experiment.

Before the start of the data assimilation, the sensitivity of the parameters was assessed in order to define the most applicable parameters to be optimized which have been set as unknown parameters. The overall methodology of this study therefore involves the development of data assimilation technique for BIOME-BGC to find the optimized ecophysiological parameters of the Teak forest. Basically, the reference data used to compare with the model output should be obtained from satellite. Thus, LAI from satellite (SPOT-LAI) was used as reference, and for this study these were considered as the observed data.

The unknown data for the simulation were the ecophysiological parameters for deciduous forest type based on Waring and Running (1998). The initial values were generated from the minimum and maximum values of the unknown parameters. GA is an optimizer that could match the observed with the simulated data by using the fitness function explained in Equation (3), until the difference of the simulated and observed data is minimized.

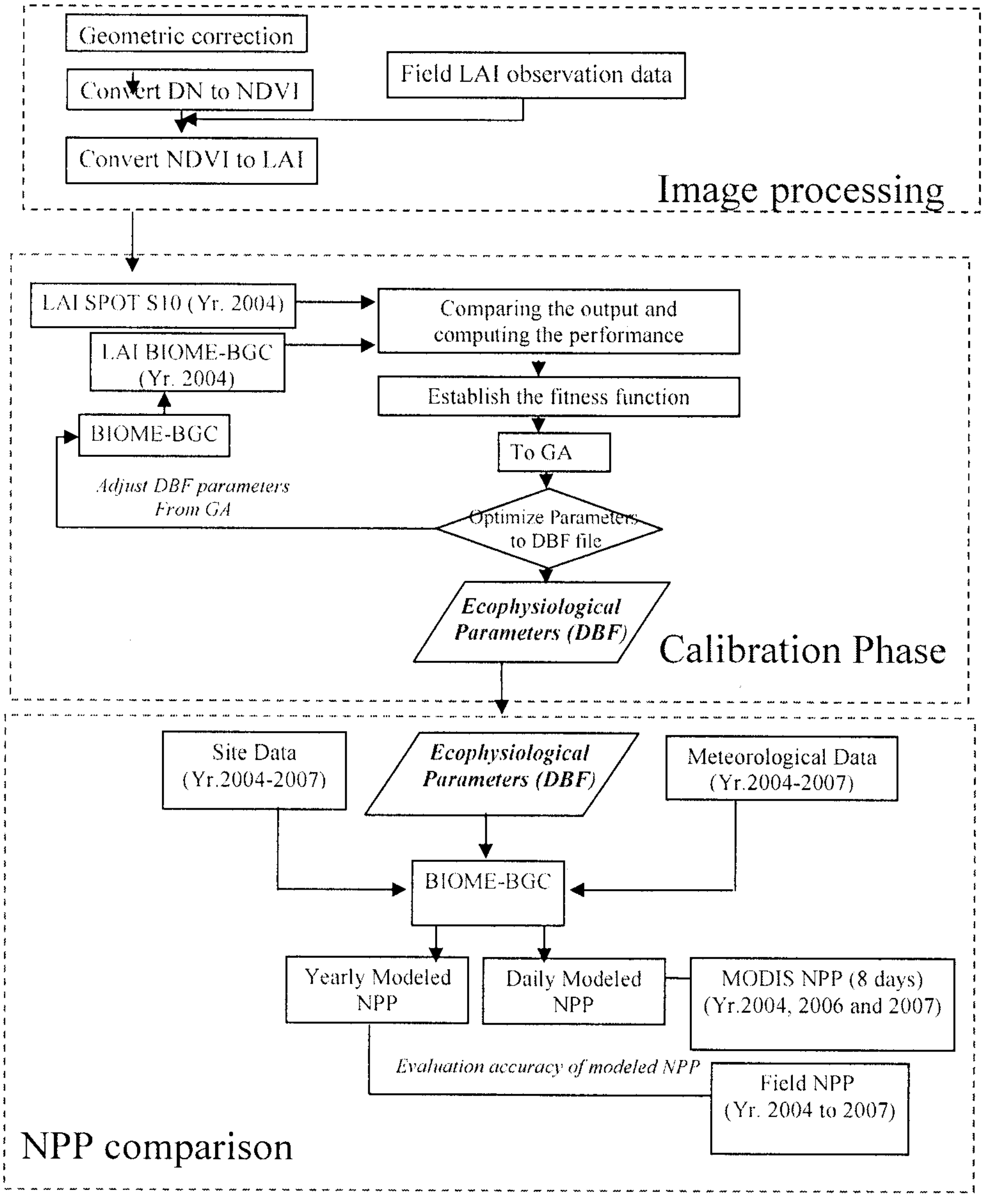


Figure 3.1 Block diagram of workflow over all methodology

NPP comparison

After the set of unknown parameters was obtained, such parameters were fed to the BIOME-BGC model along with the weather and site data to simulate the NPP. The resulting NPP data were compared with the direct field observations and with the MODIS NPP. The framework of study is shown in Figure 3.1. One limitation of the NPP validation was the inadequate and incomplete 4 year-referenced data not only from the field but also from the MODIS satellite data. Furthermore, the field data available were for year the period from 2004 to 2005 comprising the DBH and height data. The figure below also demonstrates the method for the calculation of NPP in field site and MODIS data, which were available only for the year 2004, 2006 and 2007, because the data for 2005 was incomplete. The annual NPP field data available were for year the period from 2004 to 2007.

3.2 Materials

3.2.1 Model input data

The three major groups of parameters required for the model included:

- site parameters,
- meteorological data, and
- ecophysiological characteristics of specific type of forest.

Site Parameters

The key site parameters consisting of latitude and longitude, site elevation, and soil parameter data were used for the initialization of the data file. Land Development Department (1973) indicated that the soil type in the study site is Loamy Paleustults. Typically, this type of soil has 32, 18, 50% proportion of sand, silt and clay, respectively, and was used as the soil parameter for the model. This type of soil is well drained, with medium to low organic matter and phosphorus contents as well as low potassium content. It is for these reasons that such type of soil is considered as low-fertile soil. The parameters for this type of soil are shown in Table 3.1

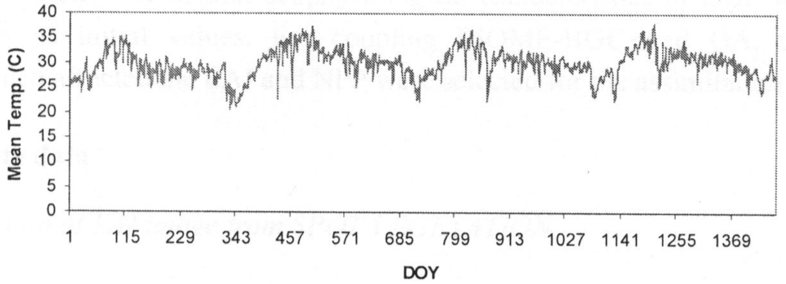
Table 3.1 Site data.

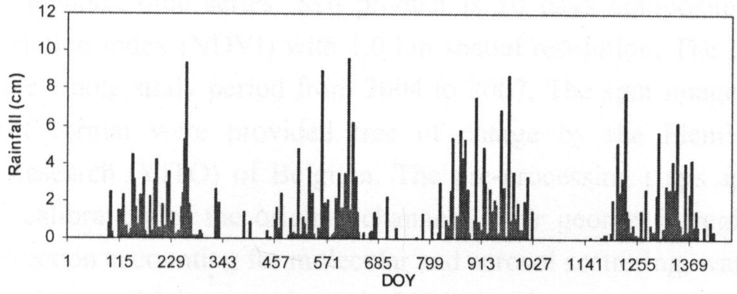
Parameter	Value	Reference
Site and soil		
Elevation (m)	380	GPS, Ground observation
Latitude (°)	18.25	GPS, Ground observation
Albedo (%)	10.6	Pinker et.al (1980)
Soil depth (m)	1.5	Ground observation
Sand:silt:clay ratio	32:18:50	Land Development Department (1973)
Nitrogen deposition (kg N m^{-2} $year^{-1}$)	0.0001	Ground observation
Nitrogen fixation (kg N m^{-2} $year^{-1}$)	0.0004	Ground observation

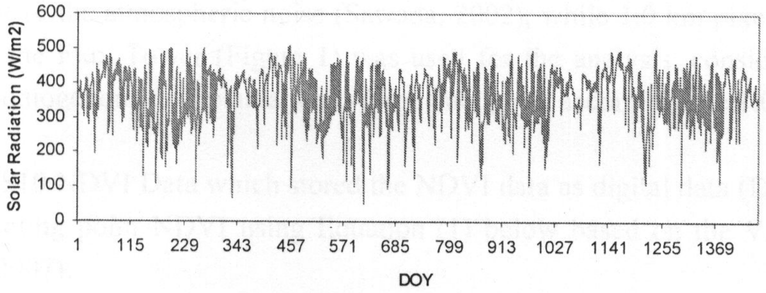
Moreover, the scenario parameters for initialization of the data include the length of simulation period, ambient CO_2 concentration (constant or variable), nitrogen deposition, among others.

Meteorological Data

BIOME-BGC determines the daily carbon and water fluxes using meteorological parameters such as precipitation, humidity, daily maximum and minimum air temperatures, mean daily air temperature, vapor pressure deficit (VPD) and incident solar radiation. Precipitation, humidity and daily maximum and minimum air temperature data were obtained from the observations at the flux tower, whereas the other parameters were obtained from MT-CLIM model which is built-in the BIOME-BGC. The measured daily maximum and minimum air temperatures were averaged using MT-CLIM to estimate the mean daily air temperature (Ta). VPD was estimated from the mean daily dew point, which is assumed to be equal to the minimum daily air temperature. The MT-CLIM model also simulates average daily incident solar radiation based on meteorological data observed at the field, as described by White et al. (2000). The meteorological inputs for BIOME-BGC are shown in Figure 3.2.







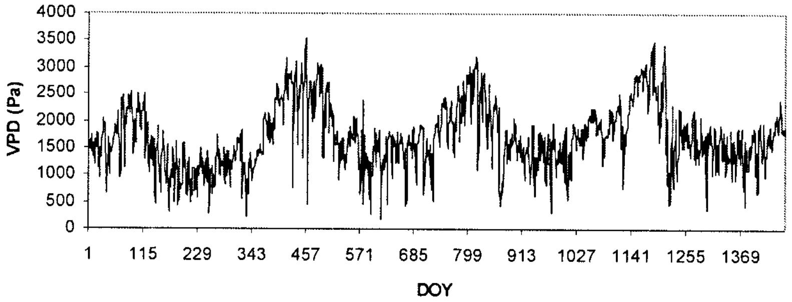


Figure 3.2 Seasonal trend meteorological input data from the year 2004-2007

Ecophysiological Characteristics

BIOME-BGC requires a static description of the ecophysiological characteristics for the type of vegetation considered in the study. A total of 43 ecophysiological parameters were defined in every biome group i.e. evergreen broadleaf forest (EBF), deciduous broadleaf forest (DBF) and etc. The default ecophysiological characteristics of DBF were used for the simulation as initial values. For coupling BIOME-BGC and GA, the 12 DBF parameters which affected the LAI and NPP were selected for the assimilation process

3.2.2 Satellite data

3.2.2.1 Derivation of LAI image from SPOT VEGETATION

The SPOT-VEGETATION S10 product was used in the assimilation of the remotely-sensed vegetation index time series. S10 product is 10 days composite of Normalized Difference Vegetation index (NDVI) with 1.0 km spatial resolution. The data used in this study covered the whole study period from 2004 to 2007. The spot images used in a pre-processed NDVI format were provided free of charge by the Flemish Institute for Technological Research (VITO) of Belgium. The pre-processing steps applied comprise the radiometric calibration of the original channels, their geometric registration and an atmospheric correction accounting for molecular and aerosol scattering, water vapor, ozone and other gas absorption (Maisongrande et al., 2004). Furthermore, the Min-Max filter was applied to remove the atmospheric noise (Sawada, 2002), while 1.0 km pixel resolution of the data from the Flux Tower (Figure 1) was used for the analysis, considering that the extent of the homogenous teak plantation is large enough to accommodate 1.0 km pixel.

The SPOT S10 NDVI Data which stored the NDVI data as digital data (DN), were then restored to floating point NDVI using Equation (1) below based on the VEGETATION Users Guide (2007):

$$NDVI = (0.004 * DN) - 0.1 \tag{3.1}$$

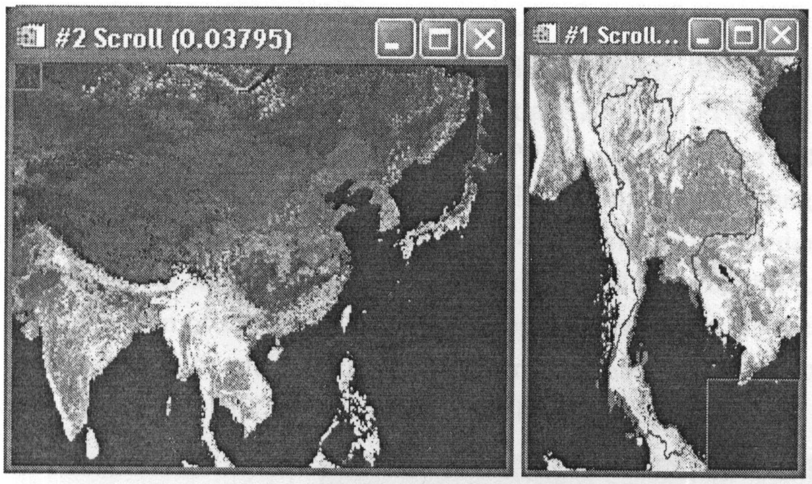


Figure 3.3 SPOT-VGT, before (left) and after (right) pre-processed subset and converted to NDVI value

For more accurate conversion of SPOT S10 NDVI to LAI data, Equation (2) which has been obtained after comparison the NDVI satellite data with LAI field observations ($R^2 = 0.84$), was applied in the study. Considering that the output of BIOME-BGC gives the absolute level of LAI, and the subsequent assimilation process compares Satellite-LAI with Simulated-LAI, the conversion of SPOT NDVI to absolute level of LAI was necessary. The formula for converting LAI from SPOT (NDVI) to LAI is indicated in Equation (2):

$$LAI = 10(NDVI) - 2.18 \tag{2}$$

After applying Equation (2), the series of LAI values after calibration were considered as reference inputs in the assimilation process. The LAI values which were retrieved by SPOT and denoted as observed data were also used in the assimilation.

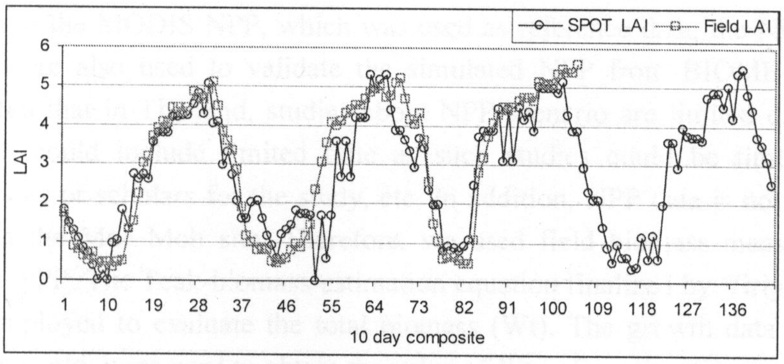


Figure 3.4 Seasonal variations in the relative LAI of sample trees, SPOT satellite LAI and Field LAI from 2004-2007

3.2.2.2 Derivation of NPP from MODIS

The referenced NPP data in this study which were derived from MODIS17A or the MODIS NPP were used to evaluate the simulated NPP from BIOME-BGC-GA. The MODIS NPP/GPP algorithm described in Running et al. (2004) and Heinsch et al. (2003), is based on the original logic of Monteith (1977), NPP is linearly related to the amount of absorbed Photosynthetically Active Radiation (PAR) during the growing period. The MODIS NPP/GPP data used in the study were the 8-day-composite data during the study period 2004-2007 at 1.0 km pixel resolution. The NPP values with the corresponding pixels within the location of the flux tower were determined according to the geographical location of the flux tower. In addition, the radiant correction and geometric correction were produced according the MODIS Swath Reprojection Tool (MRTSwath).

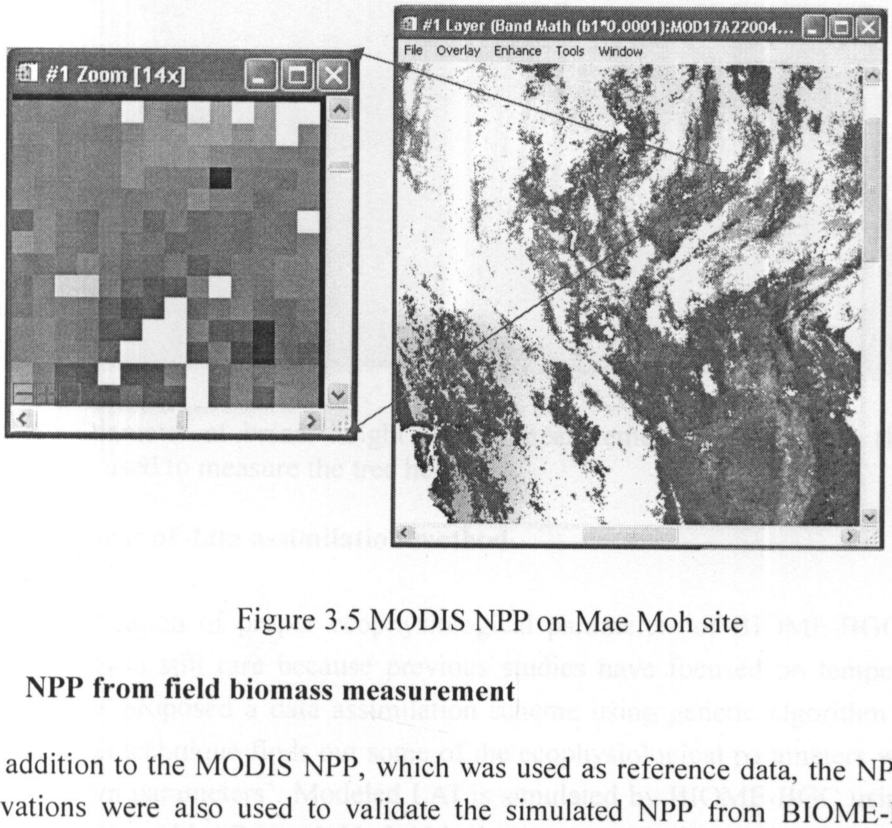


Figure 3.5 MODIS NPP on Mae Moh site

3.2.3 NPP from field biomass measurement

In addition to the MODIS NPP, which was used as reference data, the NPP from field observations were also used to validate the simulated NPP from BIOME-BGC-GA. It should be noted that in Thailand, studies about NPP scenario are limited due to various reasons which could include limited time as such studies could be time consuming, inadequate grants or scholars for the study, etc. In addition, NPP data is not available for Teak forest in the Mae Moh site. Therefore, we used field biomass measurements for estimating the NPP. The Teak-biomass-estimation equation finalized by Viriyabuncha et.al (2003) was employed to evaluate the total biomass (Wt). The growth data of teak trees aged 25-93 years old, were used to obtain the value of Wt in Equation (3)

$$Wt = 0.0166(DBH^2 * Height)^{1.0640} \qquad (R^2 = 0.99) \qquad (3.4)$$

; where Wt is Total biomass (ton/ha), DBH is diameter at breast (cm) and Height (m)

We converted the units of the following NPP estimates from dry weight of biomass to carbon by a ratio 0.475 which is the carbon content (C) proportion in biomass following the report of Raich et al. (1991).

$$NPP = 0.475 * \Delta Wt \,, \qquad (4)$$

; where NPP is Net Primary Production (kgC/m^2/yr) and ΔWt is change of biomass

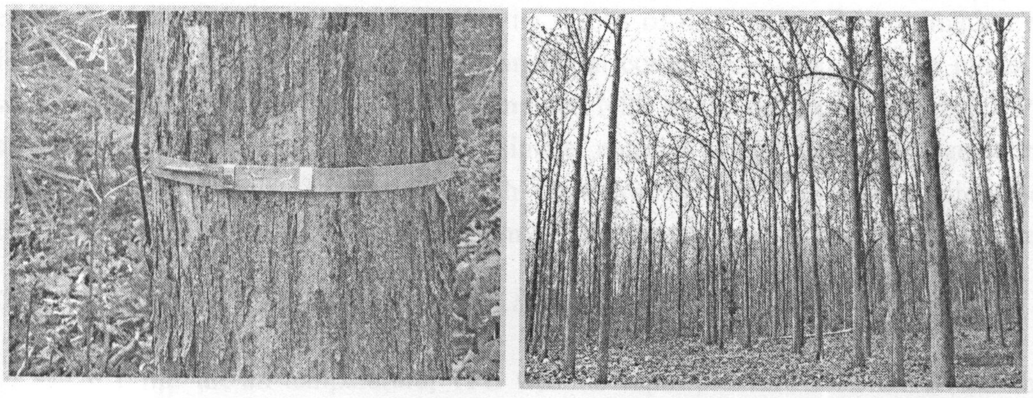


Figure 3.6 Diameter at breast height (DBH) measurement or at 1.30 m (left) Spiegel relascope was used to measure the tree height

3.3 Development of data assimilation method

The identification of proper ecophysiological parameters of BIOME-BGC model for tropical forests is still rare because previous studies have focused on temperate forests. This study has proposed a data assimilation scheme using genetic algorithm (GA) as an optimizer. This technique finds out some of the ecophysiological parameters which are set to be 'unknown parameters'. Modeled LAI is simulated by BIOME-BGC using unknown parameters proposed by GA and this LAI is then compared with observed SPOT-LAI. The difference between simulated and observed LAI is evaluated through an evaluation function called fitness function which is explained in Equation (4.1) in section 4.1. GA will change the unknown parameters and run the model again until it finds a maximum fit between the observed and simulated LAI. Through this method, it becomes possible to identify the proper BIOME-BGC model parameters which give accurate LAI as observed in field over large areas.

3.4 Implementation of the BIOME-BGC-GA model to simulate NPP

The BIOME–BGC simulates daily NPP and calculates LAI for coarsely defined biomes at areas ranging from 1 m^2 to the entire globe by using prescribed site conditions, meteorology, and parameter values. The model operates with a daily time step which means that each flux is estimated for a one-day period and weather is the most important control on vegetation processes. Leaf area index (LAI, m^2 leaf area per m^2 ground area) controls canopy radiation absorption, water interception, photosynthesis, and litter inputs to detrital pools. NPP is partitioned into the leaves, roots and stems as a function of dynamical allocation patterns, considering eventual limitations due to availability of and competition for nitrogen. NPP was calculated in terms of gas exchange, as a difference between GPP and autotrophic respiration (Ra). The BIOME-BGC method for NPP estimation is vital key to understanding the results of this study. This estimation results from the interactions of numerous environmental controls simulated by the model (as it was mentioned earlier, BIOME-BGC computed NPP as the difference between GPP and autotrophic respiration). Consequently, climate, nutrient availability, and vegetation type influence NPP through controls on both photosynthesis and respiration processes. In BIOME-BGC, the gross photosynthesis limited by climate and nutrients was calculated as:

$GPP = f(T, VPD, SW, SRAD, CO2, LAI, LEAFN)$,

where T was the air temperature, VPD was the vapor pressure deficit, SW was the soil water content, SRAD was the solar radiation at the top of canopy, CO2 was the carbon dioxide concentration in the atmosphere, LAI was the leaf area index, and LEAFN was the nitrogen concentration of leaves. Air temperature, leaf, and root nitrogen contents controlled autotrophic respiration:

$Ra = f(T, LEAFN, ROOTN)$,

where ROOTN was the nitrogen concentration of roots. Thus, BIOME-BGC was able to capture effects of a number of abiotic (temperature, vapor pressure deficit, soil water, solar radiation, and CO2 concentration) and biotic (leaf area index, leaf, and root nitrogen contents) control on NPP.

The BIOME-BGC model version 4.1.1 was used. It requires meteorological input data such as daily minimum and maximum temperature, incident solar radiation, vapor pressure deficit and precipitation. Furthermore, aspect, elevation, nitrogen deposition and fixation, and physical soil properties are needed to calculate daily canopy interception, evaporation and transpiration; soil evaporation, outflow, water potential and water content; LAI; stomatal conductance and assimilation of sunlit and shaded canopy fractions; growth and maintenance respiration; GPP and NPP; allocation; litter-fall and decomposition; mineralization and leaching.

Ecosystem simulation models need initial values of state variables to assess the physiology, biochemistry, structure and allocation patterns of vegetation functional types or biomes. Within the self-initialization process or spin up run, the development of soil and plant carbon as well as the soil and plant nitrogen pools are modelled until the annual production, respiration, decomposition, etc. are at a steady state. Soil carbon, including organic carbon within and outside living structures, was chosen to stop a spin up run since it is the last among the carbon pools to reach a steady state within an undisturbed forest ecosystem. The input needed to run this procedure are daily climate records for a given site and the leaf carbon pool size which is set to 0.001 kg m^{-2} (White et al., 2000).

The self-initialization process or spin up run can be considered as the development of natural vegetation within an undisturbed forest ecosystem. In this context, it is important to consider that these stands were heavily managed over the past 600 years (Güde, 1960). Harvesting, grazing, litter ranking and planting of trees have affected the existing forest vegetation which may have resulted in a loss of nutrients and carbon, particularly within the soil and litter layer.

3.4.1 Model data sources: site, soil and meteorological parameters

BIOME-BGC uses daily maximum and minimum air temperatures, humidity, incident solar radiation and precipitation to determine daily carbon and water fluxes. Average daily incident shortwave radiation (Qi) was simulated using MT-CLIM logic described by white et al. (2000) base on meteorological data observed at the field. Average daily net solar radiation (Qn) was estimated using a prescribed, constant albedo for vegetation. Qn was attenuated through the vegetation canopy using Beer's formulation and a prescribed extinction coefficient modulated by LAI to derive the amount of solar radiation transmitted through the canopy (Qt). The amount of solar radiation absorbed by the canopy (Qa) was estimated as the difference between Qi and Qt. Photosynthetic photon flux density (PPFD) was estimated based on the assumption that photosynthetically active radiation represents approximately 50% of Qa (Running and Coughlan 1988)

Mean daily air temperature (Ta) was estimated as the average of the measured daily maximum and minimum air temperatures. Minimum daily air temperature was assumed equal to the mean daily dew point and was used to estimate the mean daily vapor pressure deficit (VPD). Daily soil temperatures at a 30 cm soil depth (Tsoil) were estimated using an 11 day running average of Ta (Zheng et al. 1993). Soil water potential (PSI) was estimated from soil water content, soil depth and texture information following Cosby et al. (1984). Ta, VPD, PPFD and PSI were used to estimate canopy stomatal conductance (gc) and GPP following Farquhar and von Caemmerer (1980), respectively. Ta and Tsoil were used to estimate Rm (maintenance respiration) while Tsoil and PSI were used to estimate Rh (heterotrophic respiration) (Running and Coughlan 1988).

3.4.2 Model data sources: plant ecophysiological Characteristic

Biome-BGC requires a static description of the ecophysiological characteristics of the vegetation which is Teak (Tectona grandis Linn. F.) in the study area, then deciduous broadleaf group (DBF) was selected for this simulation. The general ecophysiological parameterization defined by White et al. (2000) was used, except the dates of leaf onset and offset which was particularly specified for this case.

The information on for the above-mentioned species was not available, the ecophysiological parameters for default values for deciduous broadleaved forests (DBF) were used based on White et al. (2000) (Table 3.2). The range of parameter values reported by White et al. (2000) was also used here as limits for specific parameters. After develop the program for BIOME-BGC-GA linkage. The optimized parameters values were applied for specific teak parameters. The final parameter sets applied for the studied specie is presented section 4.4.

3.5 Concluding Remarks

The methodologies developed in this chapter were implemented to explore some solution to the problems states in Chapter I. The BIOME-BGC-GA linkage was applied to simulate net primary production (NPP) in tropical deciduous forest via unknown parameter setting which currently not includes complete parameter setting in the version of the model. The results of these applications are present and discussed in Chapter IV. The Program codes are all written in Visual C.

CHAPTER IV

REMOTE SESING DATA ASSIMILATION USING BIOME-BGC

This chapter, the results are addressed in 2 parts. First, the development of data assimilation technique for BIOME-BGC using remote sensing data was explained and GA performance was discussed which obtained from program development of BIOME-BGC-GA. Second, the optimized ecophysiological parameters of a deciduous broadleaf forest (Teak forest) from BIOME-BGC-GA were explained by evaluation of the calibrated model.

4.1 Development of Data Assimilation Technique

4.1.1 Selection of unknown parameters

As a rule in data assimilation for estimating parameter values, the balance between the number of control variables and quantity of data used to constrain them, is very important. Since this study did not have enough data to treat all the 43 ecophysiological parameters as control variables, the effects of the ecophysiological parameters were assessed based on a set of key output variables and were selected to be constrained for the data assimilation development. The key output variables included in the sensitivity analysis to select the unknown parameters, were the LAI and NPP.

The sensitivity analysis was focused on the effect of the ecophysiological parameters which were developed as species and location specific, and used to determine the BIOME-BGC. Other parameters such as site and climate from field-based measurements we also used exactly as observed.

Sensitivity is the effect of input parameter x on the output variable y, and is calculated as the ratio of the change of the output variable to the change of the input parameter ($\Delta y/\Delta x$), expressed in percentage. A positive sensitivity ratio means the output variable increases with increasing input parameter value and a negative sensitivity ratio implies that the output variable decreases with increasing input parameter value. In order to select the suitable parameters for the assimilation process, the parameters were ranked in terms of their effect on the modeled variable depending on the absolute value of the sensitivity ($|\Delta y/\Delta x|$ ratio. The parameters were assigned the following three categories:

parameters with a strong effect, $|\Delta y/\Delta x|$ is larger than 0.2,
parameters with a medium effect, $|\Delta y/\Delta x|$ is between 0.1 and 0.2
parameters with low effect, $|\Delta y/\Delta x|$ is less than 0.1.

The values of the ecophysiological parameters with the certain degrees of sensitivity with respect to the LAI and NPP are shown in Table 4.1. For this study, only the parameters with strong and medium effects on the output variables were considered.

Table 4.1 Sensitivity of Leaf area index (LAI) and net primary production (NPP) to single ecophysiological parameters (The bottom index indicates the ranking of the sensitivity (0.1–0.2: index 1, medium sensitivity; above 0.2: index 2, high sensitivity).

Index	Parameter Under study	Default Value	Effect (% of valiable change to % of parameter change)	
			Maximum LAI	NPP KgC/m2/yr
P6	Transfer growth period as fraction of growing season	0.2	-0.03	-0.07
P7	Litterfall as fraction of growing season	0.2	-0.05	-0.02
P8	Annual leaf and fine root turnover fraction	1	-0.01	0.00
P9	Annual live wood turnover fraction	0.7	0.01	0.02
P10	Annual whole-plant mortality fraction	0.005	-0.01	-0.01
P11	Annual fire mortality fraction	0.0025	-0.01	0.00
P12	New fine root C : new leaf C	1	$\boldsymbol{-0.14_1}$	0.06
P13	New stem C : new leaf C	2.2	$\boldsymbol{-0.22_2}$	-0.06
P14	New live wood C : new total wood C	0.1	-0.01	-0.04
P15	New croot C : new stem C	0.23	-0.05	-0.01
P16	Current growth proportion	0.5	$\boldsymbol{0.37_2}$	0.04
P17	C:N of leaves	24	$\boldsymbol{-0.24_2}$	-0.03
P18	C:N of leaf litter, after retranslocation	49	-0.01	0.00
P19	C:N of fine roots	42	0.01	0.09
P20	C:N of live wood	50	$\boldsymbol{0.18_1}$	$\boldsymbol{0.37_2}$
P21	C:N of dead wood	442	0.00	0.00
P22	Leaf litter labile proportion	0.468	-0.05	-0.05
P23	Leaf litter cellulose proportion	0.528	-0.05	-0.05
P24	Leaf litter lignin proportion	0.204	-0.05	-0.05
P25	Fine root labile proportion	0.36	0.01	0.01
P26	Fine root cellulose proportion	0.54	0.01	0.01
P27	Fine root lignin proportion	0.3	0.01	0.01
P28	Dead wood cellulose proportion	0.912	0.00	0
P29	Dead wood lignin proportion	0.288	0.00	0
P30	Canopy water interception coefficient	0.0492	0.00	0
P31	Canopy light extinction coefficient	0.7	$\boldsymbol{-0.30_2}$	$\boldsymbol{-0.75_2}$
P32	All-sided to projected leaf area ratio	2	0.00	0.00
P33	Canopy average specific leaf area (projected area basis)	30	$\boldsymbol{0.92_2}$	$\boldsymbol{-0.17_1}$
P34	Ratio of shaded SLA:sunlit SLA	2	0.06	$\boldsymbol{0.23_2}$
P35	Fraction of leaf N in Rubisco	0.08	$\boldsymbol{0.28_2}$	$\boldsymbol{0.81_2}$
P36	Maximum stomatal conductance (projected area basis)	0.005	-0.04	-0.11
P37	Cuticular conductance (projected area basis)	0.00001	0.00	0.04
P38	Boundary layer conductance (projected area basis)	0.01	-0.02	-0.05
P39	Leaf water potential: start of conductance reduction	-0.6	0.00	0.03
P40	Leaf water potential: complete conductance reduction	-2.3	0.03	$\boldsymbol{0.32_2}$
P41	Vapor pressure deficit: start of conductance reduction	930	-0.02	-0.03
P42	Vapor pressure deficit: complete conductance reduction	4100	-0.01	-0.02

4.1.2 Model development (BIOME-BGC-GA linkage)

Genetic Algorithm was coupled with the BIOME-BGC to estimate the LAI using unknown parameters. The core of the interaction program between GA and BIOME-BGC is a C program that handles the optimization. The information of GA chromosomes was passed to the BIOME-BGC by a static file that inputs the 43 ecophysiological parameters. Of these 43 parameters, 12 were allowed to vary (Table 4.2) assuming independent prior uniform distributions across feasible ranges (upper and lower boundaries) of the parameters. The upper and lower boundaries were made available for the model which randomizes the optimized parameter between these ranges. Each parameter range has no strong information about the effective parameter values in *Tectona grandis* species. Therefore for this study, literatures that indicated some upper and lower parameter values of certain species were referred to for the possible substitution.

Table 4.2 Definition of the BIOME-BGC input parameters in the chromosome.

No.	Index	Parameters	DBF (unit)	Lower Boundary	Upper Boundary	Reference
1	P4	Yearday to start new growth (ONDAY)	yday	0	364	Day of year
2	P5	Yearday to end litterfall (OFFDAY)	yday	0	364	Day of year
3	P12	New fine root C : new leaf C	ratio	0.1	5	Chiesi et al., 2007
4	P13	New stem C : new leaf C	ratio	0.1	5	Lamberty et al.,
5	P16	Current growth proportion	prop	0	1	2005
6	P17	C:N of leaves	kgC_kgN	20	90	Mitchell et al.,
7	P20	C:N of live wood	kgC_kgN	20	90	(unpublished)
8	P31	Canopy light extinction coefficient	DIM	0	1	
9	P33	Canopy average specific leaf area	^{m2}_kgC	0.5	40	
10	P34	Ratio of shaded SLA:sunlit SLA	DIM	0	5	
11	P35	Fraction of leaf N in Rubisco	DIM	0	1	
12	P40	Leaf water potential: complete conductance reduction	MPa	-4	-0.9	

The fitness function, which is the difference between the simulated and observed LAI was used as the main basis for the evaluation of unknown input parameters for future generations, where the fittest set of input parameters survives while the weak ones fade away. The spatial distributions of SPOT-LAI and BIOME-BGC-LAI were therefore used as the search criteria. The fitness function for the modeling is expressed as follows:

$$Fitness = \frac{1}{\sum_{i=1}^{N}(LAI_{SPOT,i} - LAI_{BIOME-BGC,i})^2} \tag{4}$$

where N is number of LAI-SPOT satellite during the year.

The major function of the genetic algorithm program is to increase the generation variable and execute the assimilation process until the number of generations would equal to the variable maximum number of generations (***maxgen***) setup on BIOME-BGC-GA. The best input parameters should have survived and would be obtained when the assimilation stops after reaching the ***maxgen***. The core of such interaction program is explained in Figure 4.1

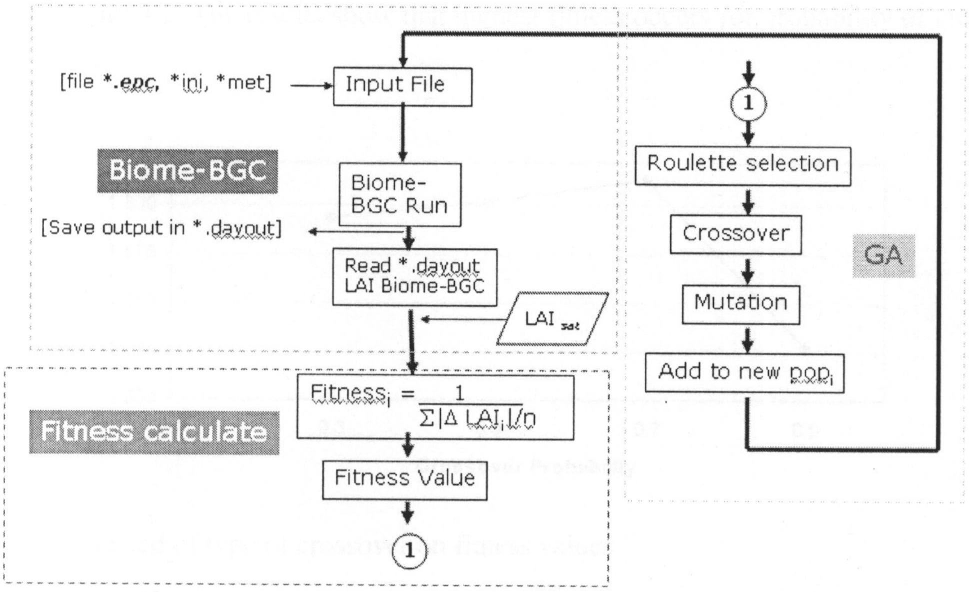


Figure 4.1 Interaction between BIOME-BGC and GA (Note: This technique would give the set of parameters that can explain the pattern and the absolute level of observation where the difference of the simulated and observed parameters is evaluated through a fitness function, for which the evaluation is repeated until GA reaches a set of parameters which gives the maximum fitness)

4.1.3 Determination of GA parameters

The determination of GA parameter has been carried out by various researchers. Goldberg (1989) found that good performance of the GA can be achieved using a high crossover probability and low mutation probability. Wardlaw and Sharif (1999) and Suiadee (2006) carried out sensitivity analysis of various parameters of GA. Wardlaw and Sharif considered crossover probabilities from 0.5 to 0.95 and recommended a crossover probability of 0.7-0.75 for real coded GA. They also recommended a mutation probability equivalent to 1 genome per chromosome such as a probability of 0.028. Suiadee found a crossover probability 0.8 as the best for his study of reservoir rule curve optimization and a mutation probability of 0.08.

GA parameters may affect the speed of convergence of results. To find the best values of the GA parameters determination testing of four different GA parameters is carried out, namely, population size, probability of crossover, mutation probability and number of generation.

Crossover Probability

This study Population size of 100 and a Pmutation = 0.09 were assumed and then investigates the fitness value for six different crossover probabilities: 0.3, 0.5, 0.7, and 0.9, as shown in Figure 4.2. The results show that highest fitness occurs for probability of crossover 0.7.

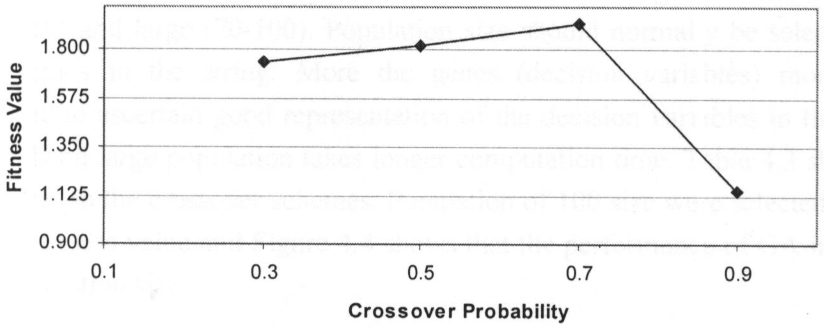


Figure 4.2 Effected of type of crossover on fitness values

Mutation Probability

Mutation probability decides the exploration in the non-visited domain through random walk. A higher value of mutation probability (e.g. larger than 50%) will result in a more explorative or random algorithm while a zero value of mutation probability will result in a chance of converging to local optima. The value of mutation probability is therefore kept considerably lower (less than 10%) than crossover probability (more than 65%) to avoid fully random search, and avoiding premature convergence. For simple problem, with a couple of decision variables, a mutation probability of 1% provides good results.

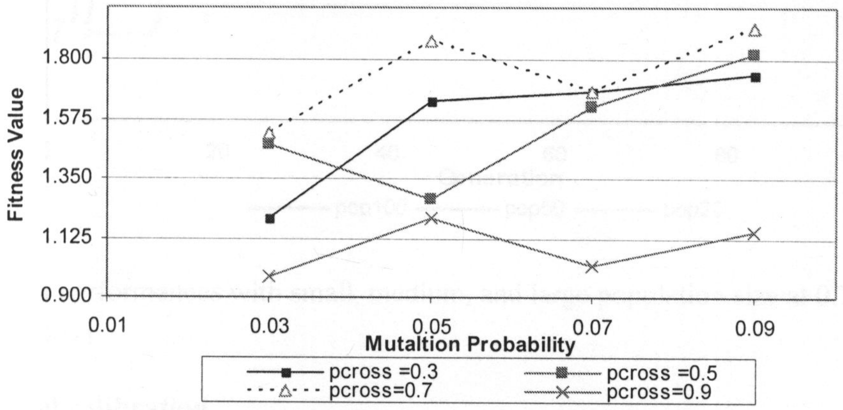


Figure 4.3 Effected of type of mutation on fitness values

The crossover probability on 0.3, 0.5, 0.7 and 0.9 were selected to test the sensitivity of mutation. The study explores the sensitivity of fitness for five mutation probabilities ranging from 0.03 to 0.09 with a step of 0.2. The maximum fitness achieved is with mutation probability of 0.09 as shown in figure 4.3

Population size

The effect of population size to the performance of GA was examined using the probability of crossover and mutation values in the previous analyses. Population sizes of 10 to 100 were used in this analysis. The population size was categorized into small (10-30), medium (40-60) and large (70-100). Population size should normally be selected based on number of genes in the string. More the genes (decision variables) more should be population size to ascertain good representation of the decision variables in the population. On the other hand large population takes longer computation time. Table 4.3 shows the best population sizes to the crossover schemes. Population of 100 size were selected in this study for more best fitness value and Figure 4.4 shows that the performance of GA improves with increasing population size.

Table 4.3 Best population size for the GA problem.

Population size	Best population
Small (10-30)	10
Medium (40-60)	50
Large (70-100)	100

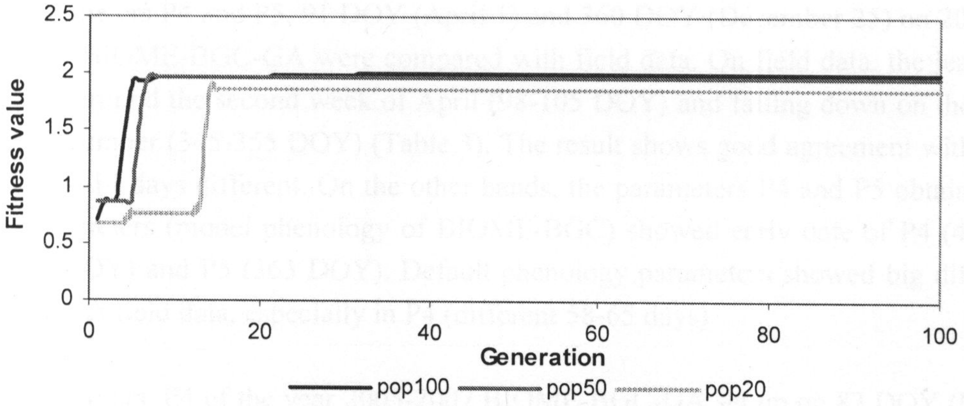


Figure 4.4 GA performances with small, medium, and large population size at 0.7 crossover and 0.09 mutation

4.2 The model calibration

The results of the model calibration could be explained in two parts. The first part shows the estimation of the parameters from assimilation compared with results of other studies, and second part is the improvement of the simulated data in the assimilation process through LAI outputs.

It is well known that the availability of accurate input parameters bridges the gap between model development and application in the field. In this study, the accuracy or improvement of the optimized parameter was evaluated through the results of the simulated LAI by using the optimized parameter and default parameter, and then observed against the satellite data which was obtained using a statistical method. The Root Mean Square Error (RMSE) is one of the most common statistics that measures the average error of a model. The lower the RMSE, the better is the performance of a model. The linear regression, R^2 measures the random error and the correspondence between the simulated and observed (satellite data). The higher correlation indicates the higher efficiency of the model.

4.2.1 Assimilation of Parameter Estimation

Initially, the 12 parameters were estimated through optimization of BIOME-BGC-GA for the year 2004. Out of the 12 parameters, the 2 phenological parameters of ONDAY and OFFDAY (P4 and P5), i.e. the start and falling date of leaves are also highly dependent on the seasonal change in the precipitation. This is normal characteristic of DBF species in tropical monsoons. The P4 and P5 then were calibrated for each year on 2004-2007 following their meteorological input data. The other 10 unknowns were calibrated only for the initial year 2004 and set to be constants value for 2005 to 2007.

The parameters P4 and P5, 91 DOY (April 1) and 360 DOY (December 25) on 2004, as obtained by BIOME-BGC-GA were compared with field data. On field data, the leaves is starting out around the second week of April (98-105 DOY) and falling down on the third week of December (345-355 DOY) (Table 3). The result shows good agreement with field value with 7-14 days different. On the other hands, the parameters P4 and P5 obtained by default parameters (model phenology of BIOME-BGC) showed early date of P4 (40, 46, 44 and 48 DOY) and P5 (363 DOY). Default phenology parameters showed big different compared with field data, especially in P4 (different 58-65 days).

For other years, P4 of the year 2005-2007 BIOME-BGC-GA set up on 83 DOY (March 23), 76 DOY (March 16) and 93 DOY (3 April), respectively. The P5 of 2005-2007, BIOME-BGC-GA set up on 361, 360 and 363 DOY respectively. P4 and P5 of 2005-2007 did not compared with the field data because of unavailable data on field.

Table 4.4 Start (ONDAY) and end (OFFDAY) of growing season

Year		P4 (ONDAY)	P5 (OFFDAY)
		DOY	DOY
2004	Field	98-105	345-355
	Optimized	91	360
	Default	40	363
2005	Optimized	83	361
	Default	46	363
2006	Optimized	76	360
	Default	44	363
2007	Optimized	93	363
	Default	48	363

The estimates of other 10 unknown parameters from the assimilation can also be considered to have similar accuracy as there is no way to estimate them in field. Further comparison with data on other tropical broadleaf forests is desirable but impossible because of unavailability of any data. However it was found that data on broadleaf forests in temperate zones are widely available [12,18,38,39] and an attempt has been made for a result comparison as shown in Table 4.5.

Table 4.5 Simulation of ecophysiological parameters by forest species on DBF (*Tectona grandis* (studied), *Populas, Quercus cerris* and Canadian DBF).

Index	Keyword	Unit	Default [9]	**STUDY RESYLT** *Tectona grandis*	*Populas* [22]	*Quercus cerris [42]*	Canadian DBF [43]
P12	New fine root C : new leaf C	Ratio	1	**0.6**	1.2	1.2	1.5
P13	New stem C : new leaf C	Ratio	2.2	**4.3**	-	2.2	2.0
P16	Current growth proportion	Prop	0.5	**0.5**	-	0.5	-
P17	C:N of leaves	kgC_kgN	24	**27**	25	25	19.8
P20	C:N of live wood	kgC_kgN	50	**51**	-	48	92
P31	Canopy light extinction coefficient	DIM	0.7	**0.9**	-	0.5	0.5
P33	Canopy average specific leaf area	m2_kgC	30	**28**	-	32	38.7
P34	Ratio of shaded SLA:sunlit SLA	DIM	2	**3**	-	2	2
P35	Fraction of leaf N in Rubisco	DIM	0.08	**0.09**	0.03	0.03	0.14
P40	Leaf water potential: complete conductance reduction	MPa	-2.3	**-1**	-	-2.2	-2.3

* *The sign (-) in table 4 means the study used default data or did not mention about that*

4.2.2 Improvement of Simulated Data in Assimilation Process

After the GA process was completed for all data from the year 2004, the best 12 ecophysiological parameters were obtained, which were then fed to the BIOME-BGC model to obtain the simulated LAI. The results showed that the BIOME-BGC-GA simulated LAI appeared to match well with the satellite LAI for the DBF vegetation types in *Tectona grandis* species. It showed a promising fit with the base values in the wet season period from the 11th to the 36th image as shown in Figure 4.3. The results showed that the BIOME-BGC-GA simulated LAI matched well with the satellite LAI for the DBF vegetation types in *Tectona grandis* species. It showed a promising fit with the base values in the wet season period from the 1 April 2004 (91 DOY) to 25 Dec 2004 (360 DOY) images as shown in Figure 4.5. The result showed that the accuracy of the estimate LAI can be enhanced by improving the accuracy of the model parameters.

Thus, the unified assimilation operation of BIOME-BGC and GA was used to obtain the optimally estimated model parameters. The result showed that the accuracy of the estimate can be enhanced by improving the accuracy of the model parameters. The output simulated LAI obtained from the default values for DBF of the BIOME-BGC model were compared with the satellite LAI, and the coefficient of determination (R^2= 0.57, RMSE= 1.40 m2/m2) was obtained.

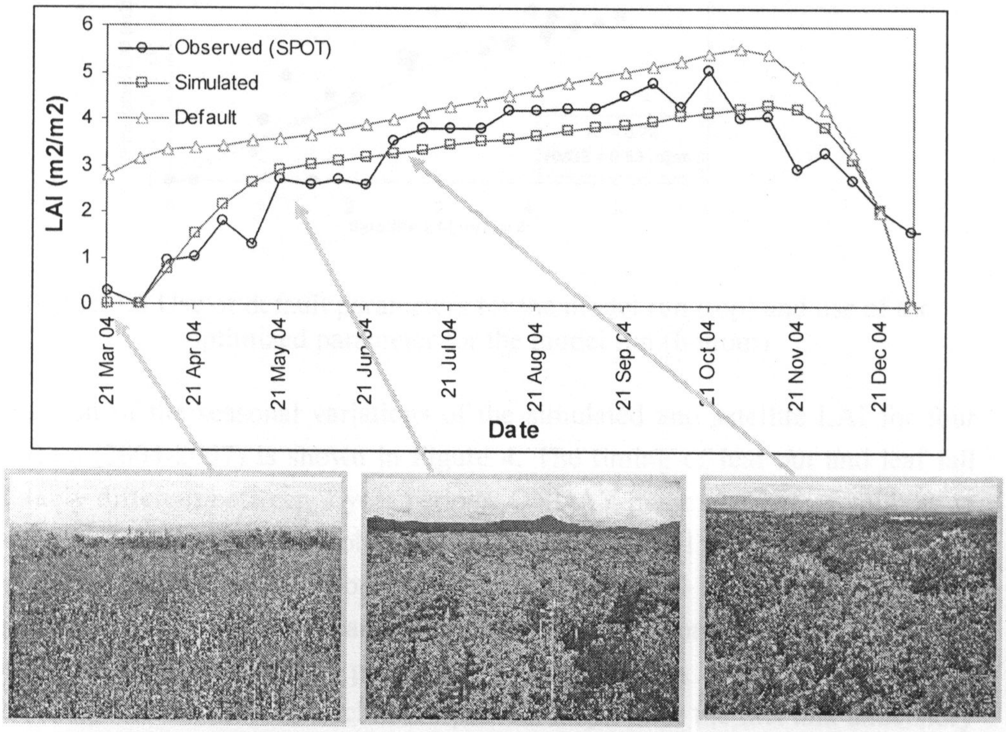
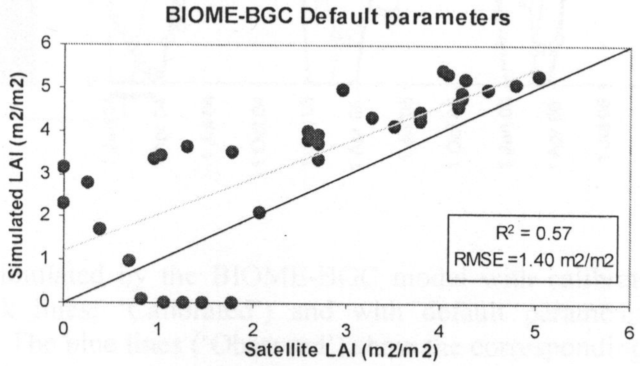


Figure 4.5 The calibration results of the BIOME-BGC-GA model (2004).

Similarly, the output LAI obtained from the GA optimized parameters for DBF of the BIOME-BGC model were also compared with the satellite LAI. This time a much improved coefficient of determination (R^2=0.78, RMSE= 0.53 m2/m2) was obtained. The result of the improved LAI model output from the optimized parameters showed appropriateness for application in the optimization of the ecophysiological values for the *Tectona grandis* species. The graphs of the regression analysis for these two scenarios can be compared, as shown in Figure 4.6.



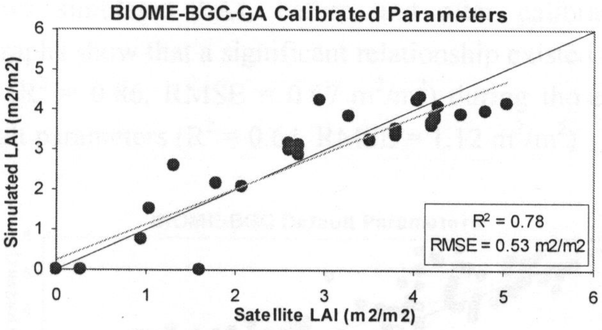


Figure 4.6 Use of default parameters for the model run (top) and use of the optimized parameter for the model run (bottom)

The comparison of the seasonal variations of the simulated and satellite LAI for four consecutive years (2004-2007) is shown in Figure 4. The timing of leaf out and leaf fall were considerably different between 4 year periods. ONDAY parameter before calibration shows an earlier leaf starting date than observed and shorter leaf fall periods than observed. During the dry season period such as from the 1 January 2004 to 21 March 2004, however, some discrepancies could be observed and the simulated LAI is somewhat lower than the observed satellite values during that period. Although this was under-predicted, it is acceptable for daily LAI estimation. Such discrepancies may due to the fact that understory of forest species (i.e. *Azadirachta indica, Xylia xylocarpa, Vitex peduncularis etc.*) present in the study area (as show in the figure 4) and moreover the local species push LAI up in rainy season on RS data.

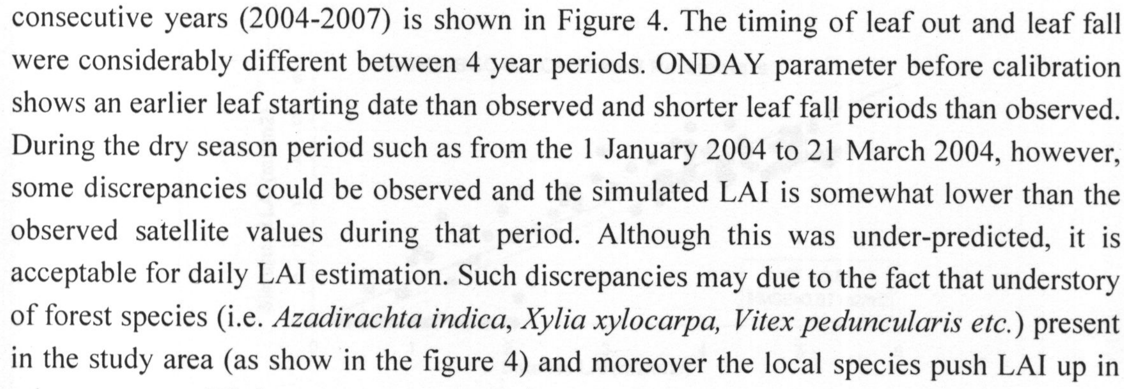


Figure 4.7 LAI simulated by the BIOME-BGC model with calibrated parameter (solid back lines; 'Calibrated') and with default parameter (soft lines; 'Default'). The blue lines ('Observed') show the corresponding observed by SPOT satellite.

Figure 4.8 shows simulated LAI, before and after calibration, compared with observations. The graphs show that a significant relationship existed between the simulated and measured LAI ($R^2 = 0.86$, RMSE = 0.67 m^2/m^2) during those years which appears better than the default parameters ($R^2 = 0.64$, RMSE = 1.12 m^2/m^2).

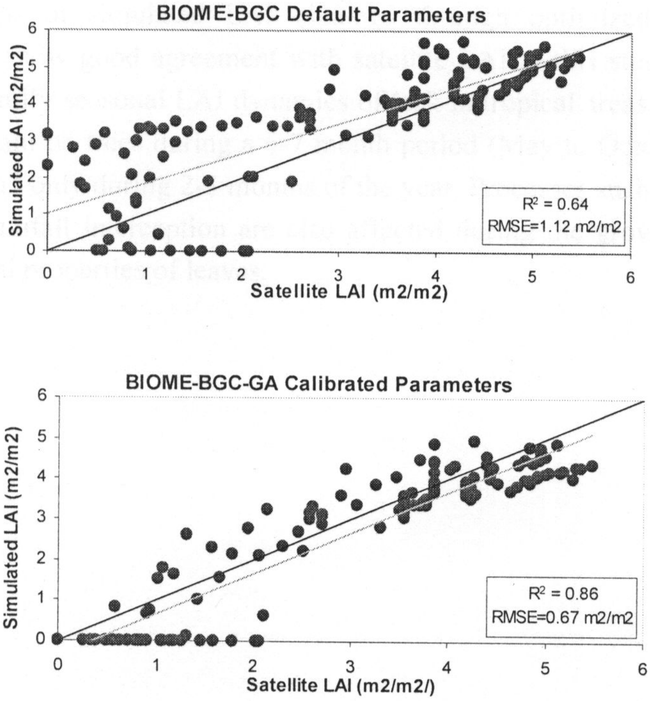


Figure 4.8 Inter-annual variation of LAI from 2004-2007: Model simulated LAI before calibration (left) compared with model simulated LAI after calibration (right)

4.3 Concluding Remarks

This chapter describes the results of the development data assimilation technique for BIOME-BGC using remote sensing (first objective) data. The determination of the GA parameters revealed important aspects for fixing the value of the GA parameters. The power of GA has been explored in testing the hypothesis postulated in Chapter 1.

Following second objective, finalized modified parameters were obtained for BIOME-BGC model for the study site at Mae Moh for teak plantation. The first two of ten ecophysiological parameters, P4 and P5 or the start and ending leaves in DBF type have evolved to acclimate to the seasonal dry period by minimizing their water use when compared with the rainy pattern in Table 3. The study showed leaf-out around the date 91 (April) of year and leaf-fall around the date 356 (December) of year (Table 4.4) used to specify in BIOME-BGC model. Some species are early and delayed for leaf out and leaf-fall which require to specify in the model as in the *Poplar* in northern Italy identify P4 was

78 day of year and P5 was 315 day of year (Migliavacca et, al., 2009) and Quercus cerris in Mediterranean forest identify P4 was 100 day of year and P5 was 300 day of year (Chiesia et,al., 2007). Other modified parameters are appropriate for apply for teak when seeing the pattern of LAI model output compared with LAI satellite.

The comparisons of simulated LAI obtained through optimized parameters from BIOME-BGC-GA show good agreement with satellite LAI at this study site. Our model can be used to quantify seasonal LAI dynamics of teak in tropical areas. At Mae Moh site, leaves usually remain on trees during a 6-7 month period (May to October), but LAIs are higher than 4 m2/m2 only during 2-3 months of the year. Processes such as photosynthesis, gas exchange or rainfall interception are also affected during the growing season by the change in functional properties of leaves.

CHAPTER V

NET PRIMARY PRODUCTION ESTIMATION

5.1 Simulation daily NPP

The BIOME-BGC model was validated daily using MODIS NPP. As shown in Figure 5.1, results of the comparison between the modeled NPP from optimized parameters (below figure) and the MODIS NPP, revealed an acceptable RMSE at 0.001 $kgC/m^2/d$ which is better than the model NPP from default parameters (top figure) with RMSE at 0.002 $kgC/m^2/d$. The improved results denote that the identification of the correct parameters was reliable.

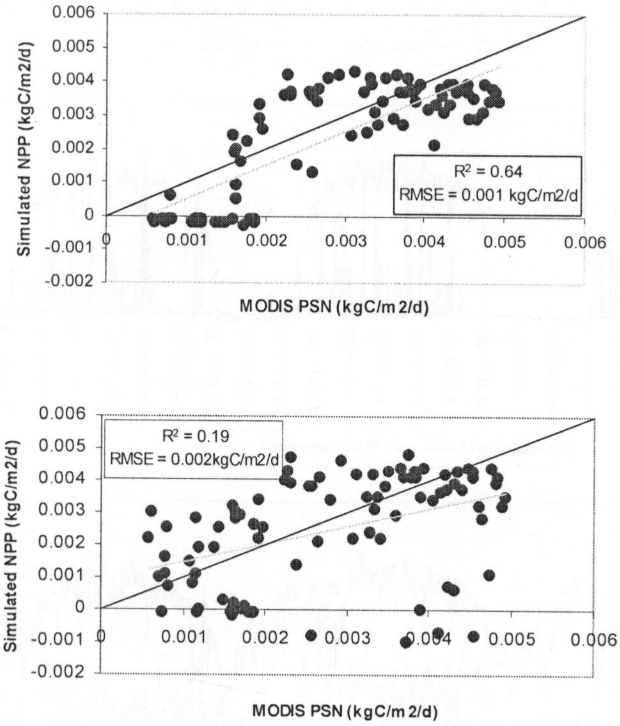


Figure 5.1 Correlation between the model NPP and MODIS NPP.

The Modeled NPP resulting from the use of the optimized parameters (Figure 5.2), illustrated the negative NPP value (-0.001 $kgC/m^2/d$) between January to beginning of April between the year 2004 and 2006-2007. The negative NPP was caused by the period of the vegetation's onset and offset (P4 and P5) on the model which was particularly specific for this site. The negative NPP indicates that plant respiration is greater than the uptake of carbon by plants during a day when vegetation is stressed by drought conditions, high VPD or low temperatures (cool dry). The BIOME-BGC estimated that the highest daily NPP occurred during the rainy period with the peak occuring in August-September. The highest teak NPP ranged from 0.003 $kgC/m^2/d$ to 0.004 $kgC/m^2/d$.

The MOIDS presented the positive value of NPP on the same period (0.004-0.005 kgC/m^2/d). Such value was still not largely different from the value of the BIOME-BGC. Therefore, their values followed the same trend. Moreover, there was also good agreement or increased trend of the NPP estimating the peak during rainy seasons. It should be noted however that due to the lack of complete data on the MODIS NPP in 2005, the data for that year was not evaluated.

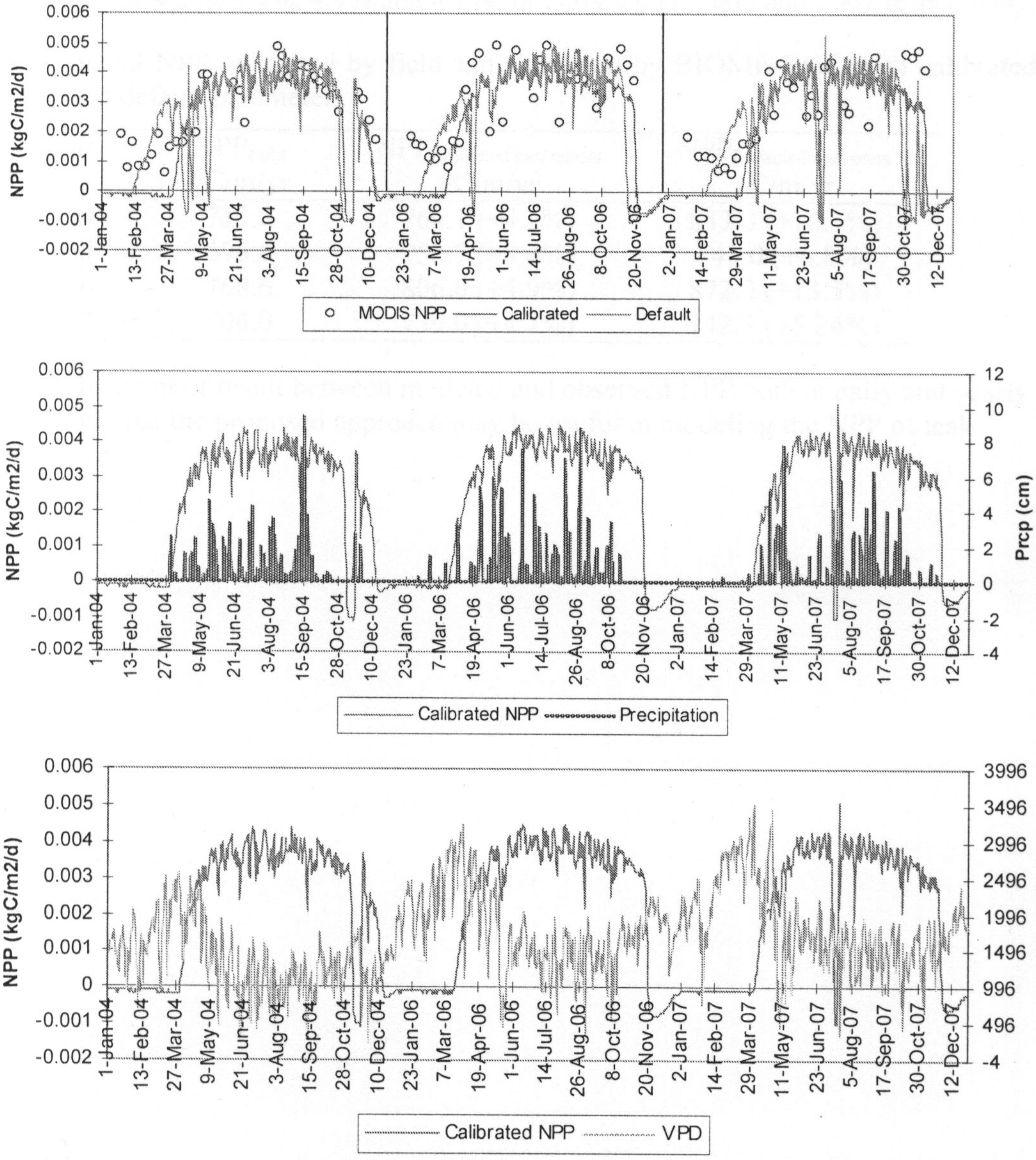


Figure 5.2 Comparison between the NPP derived from this model and NPP derived from MODIS, and Calibrated NPP with meteorological parameters.

5.2 Simulation annual NPP

The effect of correct ecophysiological parameters on the determination of the annual NPP budget is reported in Table 6 where the yearly NPP estimated with the field measured NPP and modeled NPP. The estimated NPP from modeled BIOM-BGC-GA of the four-year evaluation period were 764.1, 752.7, 806.0 and 739.0 gC/m^2/y, respectively. The modeled NPP of the 36-39 year-old trees in Mae Moh with default parameters is higher than the measured NPP (641.3-706.0 gC/m^2/y). Sensible overestimations of the yearly NPP (16.3% for the year 2004, 6.8% for the year 2005, 13.5% for the year 2006 and 5.24% for the year 2007) were then compared with the measure data. The yearly NPP estimated using the BIOME-BGC-GA with optimized ecophysiological parameters showed good accuracy with overestimate 4.2%, 8.4%, 4.9% and 4.7% for 2004, 2005, 2006 and 2007 respectively.

Table 5.1 Annual NPP measured by field and simulated by BIOME-BGC with calibrated parameters and default parameters

Year	NPP_{field} gC/m/yr	$NPP_{calibrated\ parameters}$ gC/m/yr	$NPP_{Default-Parameters}$ gC/m/yr
2004	733.0	764.1 (+4.2%)	853.0 (+16.3%)
2005	696.0	752.7 (+8.4%)	743.0 (+6.8%)
2006	768.6	806.0 (+4.9%)	872.0 (+13.5%)
2007	706.0	739.0 (+4.7%)	742.8 (+5.24%)

The good agreement of result between modeled and observed NPP both at daily and yearly can underscore that the proposed approach may be useful in modeling the NPP of teak species

CHAPTER VI

CONCLUSTIONS AND RECOMMENDATIONS

6.1 Conclusions

It is well known that the availability of accurate input parameters bridges the gap between model development and application in the field. In this paper, a modeling study conducted at site level represented the first step towards the analysis of the carbon budget of DBF of teak at a large scale, and development of the BIOME-BGC coupling the GA.

From this research, a set of relevant ecophysiological parameters which are well suited for the application of BIOME-BGC-GA for teak (DBF) plantations is also provided. The results indicated that the accuracy of the optimized model simulations is improved. The linear regression analysis between observed and simulated data showed an increase in the coefficient of determination (from 0.57 to 0.81) and a decrease in RMSE (from 1.14 to 0.71 m^2/m^2) between optimized value and default value.

As a validation exercise, the accuracy of the optimized model was evaluated using the NPP satellite data collected during 2004 to 2007 (data in 2005 was not available), achieving a good improvement in the NPP estimation with respect to the results obtained by the BIOME-BGC-GA using the default literature parameterization. The modeled NPP with the optimized and original parameters explained an improved amount of variance of the satellite NPP (R^2 from 0.19 to 0.64) and the RMSE decreased from 0.002 $kgC/m^2/d$ to 0.001 $kgC/m^2/d$. This improvement in the accuracy of the model underscores the fact that the main effect introduced by the optimized parameters had reduced the bias with the reduction of the systematic underestimation of the model.

Coupling of the GA with BIOME-BGC model is successfully carried out. The suitable GA parameter can solve the optimized problem in the study. This finding has a significant implication to large-scale problem. Finally, this research was concluded that using BIOME-BGC-GA to develop the ecophysiological parameters is crucial for NPP simulation playing a significant role in forest carbon study and forest management. This is suitable for various applications such as forest management and NPP monitoring. Therefore, a reliable data assimilation technique can help in obtaining the reliable model parameters to improve the estimates of LAI and NPP. This paper therefore presents a better way of monitoring the NPP via reduced cost laboratory experiments using the GA-based ecological model, which is the BIOME-BGC.

Despite the study did not calibrated parameters in local species which mix in the area. For future study can study parameters in various species in the area for more accuracy on

NPP estimation. BIOME-BGC-GA can applicable for optimized parameters in other species by study on pixel by pixel.

6.2 Recommendations

The following are recommendations for future studies:

(1) The implementation of the methodology on natural deciduous forest in tropical area (multi species). Finding ecophysiological parameters in other species of tropical deciduous forest.

(2) The implementation of the data assimilation methodology for NPP estimation in regional or whole of country scale by study on pixel by pixel.

(3) The inclusion of the sensitive parameter on site parameters (i.e. percentage of sand, silt, clay) as unknown parameter in data assimilation method. It is useful for NPP estimation on large scale that we did not know or unavailable about that data.

REFERENCES

Abd W. Rasib, Latif Ibrahim Ab., Cracknell A.P., Mohd Azahari Fandi and Wan Hazli Wan Kadir.., 2007 Mapping Net Primary Production in Tropical Rain Forest Using MODIS Satellite Data. *Proceedings of The Asian Conference on Remote Sensing was held on November 2007.*

Aber J. D., Reich P. B., and Goulden M. L., 1996, Extrapolating leaf CO_2 exchange to the canopy: A generalized model of forest photosynthesis compared with measurements 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 northern hardwood forests. *Global Biogeochemical Cycles 7,634-648*

Abha C. and V.K. Dadhwal., 2005, Estimating spatical, seasonal and temporal variability in terrestrial net primary productivity over India using SPOT-VEGETATION data. *Proceedings of the Second International VEGETATION User Conference 179-185.*

Akhter, S., Honda K., Chemin Y., and Uthayopas P., 2007, Exploring Strategies for Parallel Computing of RS Data Assimilation with SWAP-GA. *Journal of Computer Science. 3(1),47-50*

Asrar G., Fuchs M., Kanemasu E.T. and Hatfield J.L., 1984., Estimating absorbed photosynthetic radiation and leaf area index from spec-tral reflectance in wheat. *Agron J., 76,300-306.*

Awaya Y. and Kawahata H., 2006., Global Climate Change and Response of Carbon Cycle in the Equatorial Pacific and Indian Oceans and Adjacent Landmasses. *Elsevier. 361-377.*

Barr A.G., Black T.A., Hogg E.H., Kljun N., Morgenstern K. and Nesic Z., 2004, Inter-annual variability in the leaf area index of a boreal aspen-hazelnut forest in relation to net ecosystem production. *Agric. For. Meteorol. **126** 237-255*

Bebarta.K.C. 1999. TEAK Ecology, Silviculture, Management and Profitability. *International book distributors 46-102*

Biome-BGC: Ecophysiological parameterization, http://www.ntsg.umt.edu/ecosystem_modeling/ BiomeBGC/bgc_epc.htm

Bonan G.B., Levis S., Kergoat L. and Oleson K.W., 2002 Landscapes as patches of plant functional types: An integration concept for climate and ecosystem models. *Global Biogeochem. Cycles, **16**, doi:10.1029/2000GB001360*

Bonan, G.B., The land surface climatology of the NCAR Land Surface Model coupled to the NCAR Community Climate Model. *J. Climate, **11**, 1307-1326, 1998.*

Bond-Lamberty B., Gower S.T., Ahl D.E. and Thornton P.E., 2005, Reimplementation of the BIOME-BGC model to simulate successional change. *Tree Physiol. **25** (4), 413–424.*

Borchert R.; Rivera G., 2002, Modification of vegetative phenology in a tropical semi-deciduous forest by abnormal drought and rain. *Biotropica, **34**, 27-39*

Cramer, W., Kicklighter, D.W., Bondeau, A., Moore III, B., Churkina, G., Nemry, B., Ruimy, A. and Schloss, S.L., 1999. Comparing global models of terrestrial net primary productivity (NPP): overview and key results. Global Change Biology 5, pp. 1-15.

Cescatti A., 1997, Modelling the radiative transfer in discontinuous canopies of asymmetric crowns. II. Model testing and application in a Norway spruce stand, *Ecol. Model. **101**, 275–284*

Chapin F.S. III, Matson P.A., Mooney H.A., 2002, Principles of Terrestrial Ecosystem Ecology.New York: Springer-Verlag.

Chaurasia S., DADHWAL V.K., 2004, Comparison of principal component inversion with VI-empirical approach for LAI estimation using simulated reflectance data. *Internat. J. Remote Sensing, **25(14)**: 2881-2887*

Chemin, Y., and Honda, K., 2006, Spatio-temporal fusion of rice actual evapotranspiration with genetic algorithms and an agro-hydrological model. *IEEE T. Geosc. Rem. Sens. **44(11)**, 3462-3469*

Chemin, Y., Honda, K., and Ines, A.V.M., 2005, Genetic algorithm for Assimilating Remotely Sensed Evapotranspiration Data using a Soil-Water-Atmosphere-Plant Model Implementation Issues: *International Journal of Geoinformatics. **1(1)**,87-90. 505*

Chiesia M.; Masellia, F.; Moriondob, M.; Fibbia, L.; Bindib, M.; Running S.W., 2007, Application of BIOME-BGC to simulate Mediterranean forest processes, *Ecological Modelling, **206**, 179-190.*

Churkina G. and Running S.W., Contrasting environmental controls on the estimated productivity of different biomes. *Ecosystems **1(2)**: 206-215.*

Churkina G. and Running St., 2004, Investigating the Balance between Timber Harvest and Productivity of Global Coniferous Forests under Global Change. *Journal of Climate Change **47**,167-191*

Clark D.D., Brown S., Kicklighter D.W., Chambers J.Q., Thomlinson J.R., Ni J. and Holland, E.A. 2001. NPP in tropical forests: an evaluation and synthesis of existing field data. *Ecological Applications **11**, 371–384.*

Clark, A. D., Brown, S., Kicklighter, W. D., Chambers, Q. J. Thomlinson, R. J., Jian Ni and Holland A. E. 2001. Net primary production in tropical forests: An evaluation and synthesis of existing field data. *Ecological Applications*, 11(2). pp: 371-384.

Cramer, W. et al., 2001, Global response of terrestrial ecosystem structure and function to CO2 and climate change: results from six dynamic global vegetation models. *Global Change Biol. 4, 357–373.* (doi:10.1046/j.1365-2486.2001.00383.x)

Deborah A. C, Brown S, Kicklighter D. W., Chambers J. Q., Thomlinson J. R., Ni J, 2001. Measuring Net Primary Production in Forest: Concepts and Field Method. *Ecological Applications **11(2)**, 356-370*

Dye D.G., and Goward S.N., 1993, Photosyntheticaly active radiation by global land vegetation August 1984. *Int. J. Remote Sens., **14**,3361-3364.*

Eck T.F., and Dye D.G., 1991, Satellite estimation of Incident Photosynthetically Active radiation using ultraviolet reflectance. *Remote Sens. Environ. 38, 138-146.*

Faraway, J. and Chatfield, C., 1998, Time series forecasting with neural networks: a comparitive study using airline data. *Appl. Statistics 47 (2),231–250.*

Foley J. A., Prentice I. C., Ramankutty N., Levis S., Pollard D., Sitch S., and Haxeltine A., 1996, An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics. *Global Biogeochemical Cycles **10(4)**, 603-628.*

Furumi S., Chen L., Muramatsu K., Xiong Y., Honda Y., Fujiwara N., 2003, NPP Estimation Using Unmanned Helicopter and Satellite Observation Data on Cedar Forest in Japan. *Proceeding of ASIA GIS 2008 International Conference*

Furumi S., Muramatsu K., Ono A. and Fujiwara N., Development of estimation model for net primary production by vegetation, *Advances in Space Research **30 (11)**, 2517–2522.*

Gamo M., Panuthai S., Maeda T., Toma T., Ishida A., Hayashi M., Warsudi, Dianna R., Diloksumpun S, Phanumard L., Staporn D., Ishizuka M., Saigusa N. and Kondo H., 2005, Carbon flux observation in the tropical seasonal forests and tropical rain forest, *Proceedings of the International Workshop on Advanced Flux Network and Flux Evaluation (AsiaFlux Workshop 2005)* Fujiyoshida, Japan, 86.

Goldberg D.E., 1989, Genetic Algorithms in search and optimization and machine learning. *Addison-Wesley Publ Co Inc. USA 412.*

Goward S. N., Tucker C. J., and Dye D. G.,1985, North American vegetation patterns observed with the NOAA-7 advanced very high resolution radiometer. *Vegetation 64,3-14.*

Gower S. T., Kucharik C. J., Norman J. M., 1999, Direct and indirect estimation of leaf area index, fapar, and net primary production of terrestrial ecosystems, *Remote Sens. Environ, **70**, 29– 51.*

Gower, S.T., Krankina, O., Olson, R.J., Apps, M., Linder, S. & Wang, C., 2001, Net primary production and carbon allocation patterns of boreal forest ecosystems. *Ecological Applications **11**, 1395-1411.*

Grace, J., Malhi, Y., Higuchi, N., Meir, P., 2001, Productivity of tropical rain forests. In: Roy, J., Saugier, B., Mooney, H.A. (Eds.), *Terrestrial Global Productivity. Academic Press, New York,. 401–426*

Harley P.C., Tenhunen. J.D., Beyschlag W., Lange O.L., 1987, Seasonal changes in net photosynthesis rates and photosynthetic capacity in leaves of Cistus salvifolius, a European mediterranean semi-deciduous shrub, *Oecologia, 74, 380-388.*

Haxeltine, A., and I. C. Prentice (1996), BIOME3: An Equilibrium Terrestrial Biosphere Model Based on Ecophysiological Constraints, Resource Availability, and Competition Among Plant Functional Types, *Global Biogeochem. Cycles*, 10(4), 693–709.

Heinsch F. A., Reeves M., and Bowker C. F., 2003, User's Guide, GPP and NPP (MOD 17A2/A3) Products, NASA MODIS Land Algorithm. http://www.forestry.umt.edu/ntsg/

Hessl A.E., Milesi C., White M.A., Peterson D.L., and Keane R.E., 2004, Ecophysiological parameters for Pacific Northwest trees Gen. Tech. Rep. PNW-GTR-618. *Pacific Northwest Research Station, Portland, OR. 14.*

Hiura, T., 2005, Estimation of aboveground biomass and net biomass increment in a cool temperate forest on a landscape scale. *Ecol, Res., 20, 271-277.*

Holdridge L.R., 1967, Life Zone Ecology. San Jose, Costa Rica: Tropical Science Center. San Jose, Costa Rica

Holland H., 1975, Adaptation in Natural and Artificial Systems, *University of Michigan Press, Ann Arbor, MI.*

Hooda R.S. and Dye D.G., 1995, Identification and mapping of irrigated vegetation using.

Houghton J. T., Ding Y., Griggs D. J., Noguer M., van der Linden P. J., and Xiaosu D., Eds., 2001, Climate Change 2001: The Scientific Basis: Contributions of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. *Cambridge University Press, 881.*

Hunt E. R. Jr., 1994, Relationship between woody biomass and PAR conversion efficiency for estimating net primary production from NDVI. *Int. J.Remtore Sens.* ***15****,1725-1730.*

Hunt E. R., Piper Jr., S. C., Nemani R., Keeling C. D., Otto R. D., and Running S. W., 1996, Global net carbon exchange and intra-annual atmospheric CO_2 concentrations predicted by an ecosystem process model and three-dimensional atmospheric transport model. *Global Biogeochem. Cycles,* ***10****, 431–456.*

Ines A. V.M. and Mohanty B. P., 2008, Near-Surface Soil Moisture Assimilation for Quantifying Effective Soil Hydraulic Properties under Different Hydroclimatic Conditions. *Vadose Zone J. 7(1), 39 - 52.*

Ines A.V.M., and Honda K.; 2005, On quantifying agricultural and water management practices from low spatial resolution RS data using genetic algorithms: A numerical study for mixed pixel environment. *Advances in Water Resources* ***28***, 856-870.

IPPC, 2001, Climate change 2001. *IPPC Third Assessment Report.* Available from: http://www.grida.no/climate/ipcc_tar/.

Jihua MengBingfang WuYuemin Zhou, 2005, Monitoring Terrestrial Net Primary Productivity of China Using BIOME-BGC Model Based on Remote Sensing. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2005),****5*** *, 3105-3108*

Jolly W.M. and Running S.W., 2003, Effects of precipitation and soil water potential on drought deciduous phenology in the Kalahari. *Global Change Biology (2003) doi: 10.1046/j.1529-8817.2003.00701.x*

Kaastra, I. and Boyd, M.S., 1995, Forecasting futures trading volume using neural networks. *J. Future Markets* ***15*** *8, 953–970*

Kang S., Kimball S.J., Running S.W., 2005, Simulating effects of fire disturbance and climate change on boreal forest productivity and evapotranspiration, *Sciences of The Total Environment*, ***362***, 85-102.

Knorr W. and Kattge J., 2005, Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling, *Global Change Biology **11** 1333–1351*

Koza J. R., 1992, *Genetic* Programming: *On the Programming of Computers by Means of Natural Selection.* Cambridge, MA: The MIT Press.

Kuroda H. and Kishi M.J., 2004, A data assimilation technique applied to estimate parameters for the NEMURO marine ecosystem model, *Ecol. Modell. **172**, 69–85.*

Land Classification. Soil Interpretation Handbook for Thailand. 1973, Department of Land Development, Ministry of Agriculture and Cooperatives, Bangkok, Thailand.

Larcher W., 1983. Physiological Plant Ecology. *Springer, New York, USA.*

Lawson L.M., Spitz Y.H., Hofmann E.E. and Long R.L., 1995, A data assimilation technique applied to a predator–prey model. *Bull. Math. Biol.* **57**, *593–617.*

Lawson M.L., Spitz Y.H., Hofmann E.E. and R.B. Long. 1995, A data assimilation technique applied to a predator-prey model. *Journal of Bulletin of Mathematical Biology. **57**, 593-617.*

Luo T., Neilson R.P., Tian H., Vörösmarty C.J., Zhu H., Liu S., 2002, A model for seasonality and distribution of leaf area index of forests and its application to China. *Journal of Vegetation Science, **13**,* 817-830.

Maass, M., and Martinez-Yrizar A., 2001., NPP Tropical Forest: Chamela, Mexico, 1982-1995. 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.*

Maisongrande P., Duchemin B. and Dedieu G., 2004, Vegetation/SPOT: An operational mission for the Earth monitoring: Presentation of new standard products. *International Journal of Remote Sensing* **25 (1)**, *9–14.*

Majdi, H., Pregitzer, K., Moren, A.-S., Nylund, J.-E., and Agren, G.I., 2005, Measuring Fine Root Turnover in Forest Ecosystems, *Plant and Soil, **276**, 1-8.*

Marks R.H., 1975, Methods of assessing terrestrial productivity. *In: Lieth, H., Whittaker, R.H. (Eds.), Primary Productivity of the Biosphere. Springer, Berlin, pp. 55–118.*

Maselli F., Barbati A., Chiesi M., Chirici G. and Corona P., 2006, Use of remotely sensed and ancillary data for estimating forest gross primary productivity in Italy, *Remote Sens. Environ. **100**, 563–575.*

Matear R.J., 1995, Parameter optimization and analysis of ecosystem models using simulated annealing: a case study at Station P. *Journal of Maine System. Res. **53**, 571–607.*

Matear R.J., 1995, Parameter optimization and analysis of ecosystem models using simulated annealing: a case study at Station P. *J. Mar. Res.* **53**, *571–607.*

Medina, E., Klinge H., 1983, Productivity of tropical forests and tropical woodlands. *Encyclopaedia Plant Physio, **12D**,* **281–303**

Melillo J.M., McGuire A.D., Kicklighter D.W., Moore III B., Vorosmarty C.J., Schloss A.L., 1993, Global climate change and terrestrial net primary production. *Nature **363**, 234-240.*

Melillo JM, McGuire AD, Kicklighter DW, Moore III B, Vorosmarty CJ, Schloss AL., 1993, Global climate change and terrestrial net primary production. *Nature. 363:234-240.*

Migliavacca M., Meroni M., Busetto L., Colombo R., Zenone T., Matteucci G., Manca G., Seufert G., 2009, Modeling Gross Primary Production of Agro-Forestry Ecosystems by Assimilation of Satellite-Derived Information in a Process-Based Model. *Sensors **9**, 922-942.*

Migliavacca M., Meroni M., Busetto L., Colombo R., Zeone T., Matteucci G., Manca G. and Seufert G., 2009, Modeling Gross Primary Production of Agro-Forestry Ecosystems by Assimilation of Satellite-Derived Information in a Process-Based model. *Sensors, **9**, 922-942*

Mirsra, R., 1973. *In*: Tropical Ecology with an Emphasis on Organic Production pp.279-294. *(Eds. P.M. Golley and F.B.Golley).Mimeo*

Mitchell S.R., Beven K.J., and Freer J.E. Multiple sources of predictive uncertainty in modeled estimates of net ecosystem CO_2 exchange. *Ecological Modelling (in press).*

Monteith J.L., 1987, Climate and efficiennncy of crop production in Britain. *Phil. Tran. Royal Soc. London, B **281**,277-294.*

Monteith, J. L., 1977, Climate and the efficiency of crop production in Britain. *Philisophical Transactions of the Royal Society of London, Series B, (281) 277-294.*

Murphy, P.G., 1975, Net primary productivity in tropical terrestrial ecosystems. In: Lieth, H., Whittaker, R.H. (Eds.), *Primary Productivity of the Biosphere. Springer, Berlin, 217–231.*

Nilson T., Lang M., Kuusk A., Anniste J. and Lükk T., 1999,Forest reflectance model as an interface between satellite images and forestry databases, *Proceedings of the Conference "Remote Sensing and Forest Monitoring"* June 1–3, Rogow, Poland (1999).

Pietsch S.A., Hasenauer H. and Thornton P.E., 2005, BGC-model parameters for tree species growing in central European forests, *For. Ecol. Manage. **211**, 264–295.*

Qian Y.L., Bandaranayake W., Parton W.J., Mecham B., Harivandi M.A. and Mosier A.R., 2003, Long-term effects of clipping and nitrogen management in turfgrass on soil organic carbon and nitrogen dynamics: the CENTURY model simulation, *J. Environ. Qual. **32**, 1694–1700.*

Raich J.W., Rastetter E.B., Melillo J.M., Kicklighter D.W., Steudler P.A., Peterson B.J., Grace A.L., Moore B III., and Vörösmarty C.J., 1991, Potential net primary productivity in South America: application of a global model. *Ecol. Appl. **1** 399–429*

Raupach M.R., Rayner P.J., Barrett D.J., Defries R.S., Heimann M., Ojima D.S., Quegan S. and Schmullius C.C., 2005, Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. *Global Change Biology **11** 378–397.*

Ruimy A., Saugier B., Dedieu G., 1994, Methodology for the estimation of terrestrial net primary productivity from remotely sensed data. *J. Geophys. Res. 99, 5263-5283.*

Running S. W., and Coughlan J. C., 1988, A general model of forest ecosystem processes for regional applications. I. Hydrological balance, canopy gas exchange and primary production processes. *Ecol. Model, **42**, 125–154.*

Running S.W., and Hunt E.R., 1993, Generalization of a forest ecosystem process model for other biomes, BIOME-BGC, and an application for global-scale models. In: J.R. Ehleringer and C.B. Field, Editors, *Scaling Physiological Processes: Leaf to Globe*, Academic Press, San Diego, 141–158.

Running S.W., and Hunt E.R., 1993, Generalization of a forest ecosystem process model for other biomes, BIOME-BGC, and an application for global-scale models. *In: J.R. Ehleringer and C.B. Field, Editors, Scaling Physiological Processes: Leaf to Clobe, Academic Press, San Diego 141-158*

Running S.W., Baldocchi D., Turner D.P., Gower S.T., Bakwin P.S., and Hibbard K.A., 1999, A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS satellite data. *Remote Sensing of Environment, **70**, 108-127.*

Running S.W., Nemani R.R. and Hungerford R.D., 1987, Extrapolation of synoptic meteorological data in mountainous terrain and its use for simulating forest evaporation and photosynthesis, *Can. J. Forest. Res. **17**, 472–483*

Running S.W., Nemani R.R., Heinsch F.A., Zhao M., Reeves M. and Hashimoto H., 2004, A continuous satellite-derived measure of global terrestrial production, *BioScience **54** 547–560.*

Running, S.W., 1994, Testing FOREST-BGC ecosystem process simulations across a climatic gradient in Oregon. *Ecological Applications, **4(2)**: 238-247*

Saugier B., Roy J. and Mooney H.A., 2001, Estimations of global terrestrial productivity: Converging toward a single number. In: J. Roy, B. Saugier and H.A. Mooney, Editors, *In Terrestrial Global Productivity, Academic Press, San Diego, CA, 543–557.*

Sawada, H. and Sawada Y., 2002, Modeling of vegetation seasonal change based on high frequency observation satellite. *Environmental Information Science Papers, **12**.*

Schlesinger W.H., 1997, Biogeochemistry: an analysis of global change, 2nd edn. Academic, New York.

Shibata M., Tanaka H., Iida S., Abe S., Masaki T., Niiyama K., Nakashizuka T., 2002, Synchronized annual seed production by 16 principal tree species in a temperate deciduous forest, Japan. *Ecology **83**, 1727-1742*

Silver W.L., Thompson A.W., McGroddy M.E., 2005, Varner R.K., Dias J.D., Silva H., Crill P.M. and Kellers M., Fine root dynamics and trace gas fluxes in two lowland tropical forest soils, *Global Change Biology **11**, 290–306*

Smit, A.L., Bengough, A.G.., Engels, C., van Noordwijk, M., Pellerin, S., and van de Geijn, S.C., 2000a, *Root Methods: A Handbook, Springer-Verlag, Heidelberg, 587.*

Smit, A.L., George, E., and Groenwold, J., 2000b, Root observatino and measurements at (Tranparent) interface with soil, In Root methods: *A handbook, edited by A.L, Smit,* *A.G, Bengough, C, Engels, M, van Noordwijk, S, Pellerin & S.C, van de Geijn, 235-271, Springer-Verlag, Heidelberg.*

Spanner M.A., Pierce L.L., Running S.W., Peterson D.L., 1990, The seasonality of AVHRR data of temperate coniferous forests: relationship with leaf area index, *Remote Sens. Environ. 33, 97-112.*

Spitz et al., 1998. Y.H. Spitz, J.R. Moisan, M.R. Abbott and J.G. Richman, 1998, Data assimilation and a peragic ecosystem model: parameterization using time series observation. *Journal of Maine System, 16, 51–68.*

Spitz Y.H., Moisan J.R., Abbott M.R. and Richman J.G., 1998, Data assimilation and a peragic ecosystem model: parameterization using time series observation. *J. Mar. Syst. 16, 51–68.*

Srinuandee P., Honda K., Chemin Y., Ines A.V.M., 2004, SWAP crop model Parameter Identification using SPOT Vegetation in Suphanburi, Thailand. *Proceedings of the 26th Asian Conference on Remote Sensing (ACRS).*

Swinbank R., Keil M., Jackson D.R., Scaife A.A., 2004, Stratospheric data assimilation at the Met Office – progress and plans. In: *ECMWFWorkshop on Modelling and Assimilation for the Stratosphere and Tropopause*, 23–26 June 2003, 147–154.

Tatarinov F.A. and Cienciala E., 2006, Application of BIOME-BGC model to managed forests. 1. Sensitivity analysis, *For. Ecol. Manage.* **237**, *267–379.*

Thornton P. E., and Running S. W., 1999: An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. *Agric. For. Meteorol, 93, 211–228.*

Tripathi, S.K., Sumida, A., Shibata, H., Uemura, S., Ono, K., and Hara, T., 2005, Growth and substrate quality of fine root and soil nitrogen availability in a young Betula ermanii forest of northern Japan: Effects of the removal of understory dwarf bamboo (Sasa kurilensis), *Forest Ecol, Manag, 212, 278-290*

Turner D. P., Ritts W. D., Cohen W. B., Gower S. T., Zhao M., Running S. W.; Wofsy, S. C., Urbanski S., Dunn A. L., Munger J.W., 2003, Scaling Gross Primary Production (GPP) over boreal and deciduous forest landscapes in support of MODIS GPP product validation. *Remo Distributed Active Archive Center (ORNL DAAC) .2009. Net Primary Productivity Method.* *http://daac.ornl.gov/NPP/html_docs/npp_est.html*

Turner D.P., Ritts W.D., Cohen W.B., Gower S.T., Zhao M., Running S.W., Wofsy S.C., Urbanski, S., Dunn A., Munger, J.W., 2003, Scaling gross primary production (GPP) over boreal and deciduous forest landscapes in support of MODIS GPP product validation. *Remote Sensing of Environment 88, 256-270.*

Tziperman E., and Thacker W.C., 1989, An Optimal–Control/adjoint—equations approach to studying the oceanic general circulation. *J. Phys. Oceanogr 19, 1471–1485.*

Vallino J.J., 2000, Improving marine ecosystem models: use of data assimilation and mesocosm experiments. *J. Mar. Res.* **58**, *117–164.*

VEGETATION Users Guide, http://free.vgt.vito.be/

Vetter M., Churkina G., Jung M., Reichstein M., Zaehle S., Bondeau A., Chen Y. H., Ciais P., Feser F., Freibauer A., Geyer R., Jones C., Papale D., Tenhunen J., Tomelleri E., Trusilova K., Viovy N., and Heimann M., 2008, Analyzing the causes and spatial pattern of the European 2003 carbon flux anomaly using seven models, *Biogeosciences **5**, 561–583*

Viriyabuncha C, Sumantakul V., 2007, Studies on Net Primary Production of Teak (*Tectona grandis* L.f.) and Some Fast Growing Tree Species Plantations for Sustainable Conservation. *Silviculture Research Group, Department of National Park, Wildlife and Plant Conservation. Bangkok, THAILAND. (not published yet)*

Vogt K.A., Vogt D.J., Palmiotto P.A., Boon P., Hara J. O and H. Asbjornsen, Review of root dynamics in forest ecosystems grouped by climate, climatic forest type and species, *Plant and Soil **187** 1996*

Wang Y.-P., Baldocchi D., Leuning R., Falge E. and Vesala T., 2006, Estimating parameters in a land-surface model by applying nonlinear inversion to eddy covariance flux measurements from eight FLUXNET sites, *Global Change Biology **12**, 1–19.*

Wang Y.-P., Leuning R., Cleugh H.A. and Coppin.A. P, 2001, Parameter estimation in surface exchange models using nonlinear inversion: how many parameters can we estimate and which measurements are most useful?, *Global Change Biology 7,495–510.*

Wang, Chales A.S. Hall, Frederick N. Scatena, Ned Fetcher, Wei Wu. 2003, Modelling the spatial and temporal variability in climate and primary productivity across the Luquillo Mountains, Puerto Rical. *Forest Ecology and Management 179 69-94*

Wardlaw, R.B., and Sharif, M. 1999, Evaluation of genetic algorithms for optimal reservoir system operation. J.Water Resour. Planning. and Mgmt., (ASCE). *125, 25-33.*

Waring H.R. and Running S.W., 1998, Forest Ecosystems. Analysis at Multiples Scales (2nd ed.), *Academic Press, San Diego p. 55.*

Waring R.H. and Running S.W., 1998, *Forest Ecosystems. Analysis at multiple scales. Academic Press, San Diego, CA, USA.*

White M. A., Thornton P. E., Running S. W., and Nemani R. R., 2002, Literature-Derived Parameters for the BIOME-BGC Terrestrial Ecosystem Model. 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.*

White M.A., Thornton P.E.; Running S.W., and Nemani R.R., 2000, Parameterization and sensitivity analysis of the BIOME-BGC terrestrial ecosystem model: net primary production controls, *Earth Interact. 4 1-85(White et al., 2000; available online at http://www.ntsg.umt.edu/)*

White M.A.; Thornton P.E.; Running S.W., 1997, A continental phenology model for monitoring vegetation responses to interannual climatic variability, *Global Biogeochem. Cycl. **11***, 217-234.

White, M. A., Thornton P. E., Running S. W., and Nemani R. R., 2000, Parameterization and Sensitivity Analysis of the BIOME-BGC Terrestrial Ecosystem Model: Net Primary Production Controls. *Earth Interactions* ***4****(3):1-85.*

Whittaker R.H. and Likens G.E. , 1975, The biosphere and man. In: Leith, H. and Whittaker, R.H. (Eds.), *Primary Productivity and the Biosphere, Ecological Studies 14, Springer-Verlag, Berlin, Germany, 305-328.*

Williams M., Schwarz P.A., Law B.E., Irvine J. and Kurpius M.R., 2005, An improved analysis of forest carbon dynamics using data assimilation. *Global Change Biology* ***11****, 89–105.*

Wilson B.; Baldocchi D.D., 2000, Seasonal and interannual variability of energy fluxes over a broadleaved temperate deciduous forest in North America, *Agric. For. Meteorol.* ***100****, 1–18*

Woodward J., 1984, A Theory of Singular Causal Explanation. *Erkenntnis* ***21****, 231-262.*

Wu J., Jones K.B., Li H., and Loucks O.L. 2006, Scaling and uncertainty analysis in ecology: methods and applications. *Springer. 167-190.*

Yoshifuji N., Kumagai T., Tanaka K., Tanaka N., Komatsu H., Suzuki M., Tantasirin C., 2006, Inter-annual variation in growing season length of a tropical seasonal forest in northern Thailand. *Forest Ecology and Management, 229, 333-339.*

APPENDICS

Appendix A: Simple genetic algorithm implementation

```
/**************************************************************/
/* This is a simple genetic algorithm implementation where the */
/* evaluation function takes positive values only and the      */
/* fitness of an individual is the same as the value of the    */
/* objective function                                 */
/**************************************************************/

#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <string.h>
#include <unistd.h>
#include <sys/time.h>

#include "gaBGC_Util.h"

/****************************/
/*UNSTABLE*/
#define NVARS 43        // for number of unknows
#define POPSIZE 10 /* population size */
#define MAXGENS 100  /* max. number of generations */

#define LAI_NUM 36 // number of LAI values to compare in each file
#define LAI_PERIOD 10 // number of satellite reading cycle period in days
#define DAYSINYEAR 365 // according to 365 days in a year

#define DAY_ERR_NUM 1 // according to the number of unrecorded sat LAI in day_err array
#define RAND_NUM 12 // according to the number of random variables
#define PXOVER 0.7        /* probability of crossover */
#define PMUTATION 0.09            /* probability of mutation */

int generation;               /* current generation number */
int cur_best;                 /* best individual */

FILE *galog;                  /* an output file */

int seed = 0;
int day_err[] = {39}; // line index in of unrecored sat LAI in satLAI data file
int arr_rand[] = {4,5,12,13,16,17,20,31,33,34,35,40}; // random variables array

struct genotype /* genotype (GT), a member of the population */
{
  char  unit[NVARS][30];
  int   digit[NVARS];
  double gene[NVARS]; /* a string of variables */
  double fitness;    /* GT's fitness */
  double upper[NVARS];   /* GT's variables upper bound */
  double lower[NVARS];   /* GT's variables lower bound */
  double rfitness;   /* relative fitness */
  double cfitness;   /* cumulative fitness */
};

struct genotype constpopulation[1]; /* constant variables population */
struct genotype population[POPSIZE+1];              /* population */
struct genotype newpopulation[POPSIZE+1];           /* new population */
                                                    /* replaces the */
                                                    /* old generation */
```

```
/* Declaration of procedures used by this genetic algorithm */

double randval(double, double, int);

void initialize(void);
void evaluate(void);
void keep_the_best(void);
void elitist(void);
void selecta(void);
void crossover(void);
void Xover(int, int);
void swap(double *, double *);
void mutate(void);
void report(void);

/***************************************************************/
/* Initialization function: Initializes the values of genes    */
/* within the variables bounds. It also initializes (to zero)  */
/* all fitness values for each member of the population. It    */
/* reads upper and lower bounds of each variable from the      */
/* input file (gadata.txt). It randomly generates values       */
/* between these bounds for each gene of each genotype in the  */
/* population. The format of the input file (gadata.txt) is    */
/* var1_lower_bound var1_upper_bound                           */
/* var2_lower_bound var2_upper_bound                           */
/***************************************************************/

void initialize(void)
{
  FILE *infile;
  int i,j, dig;
  double tval, lbound, ubound;
  char var_name[30];
  char unit[30];

        clearAll();

  if((infile = fopen("dbf_unknown.txt","r"))==NULL)
  {
    fprintf(galog,"\nCannot open dbf_unknown file!\n");
    exit(1);
  }

/* Initialize the variables within the bounds */

        for (i=0;i<NVARS;i++)
        {
     fscanf(infile, "%s",&var_name);
     fscanf(infile, "%s",&unit);
     fscanf(infile, "%d",&dig);
     fscanf(infile, "%lf",&tval);
     fscanf(infile, "%lf",&lbound);
     fscanf(infile, "%lf",&ubound);

                for(j=0;j<POPSIZE;j++)
                {
      strcpy(population[j].unit[i], unit);
      population[j].digit[i]=dig;
      population[j].fitness=0;
```

```
        population[j].rfitness=0;
        population[j].cfitness=0;
        population[j].lower[i]=lbound;
        population[j].upper[i]=ubound;

        strcpy(constpopulation[0].unit[i], unit);
        constpopulation[0].digit[i]=dig;
        constpopulation[0].fitness=0;
        constpopulation[0].rfitness=0;
        constpopulation[0].cfitness=0;
        constpopulation[0].lower[i]=lbound;
        constpopulation[0].upper[i]=ubound;

        if(isRandom(i, RAND_NUM, arr_rand)==1)
        {
          population[j].gene[i]=randval(lbound, ubound, dig);
          //printf("RANDOM var %d, pop %d, val %lf\n", i, j, population[j].gene[i]);
                              }
        else
          population[j].gene[i]=tval;
                  }
          }

fclose(infile);
}

/*********************************************************/
/* Random value generator: Generates a value within bounds */
/*********************************************************/

double randval(double low, double high, int decimal)
{
         srand((unsigned)time(NULL)+seed);
  seed = (seed+1)%1000;
  //printf("seed: %d\n", seed);
         double val;
         val=((double)(rand()%1000)/1000.0)*(high-low)+low;
         return(roundNumber(decimal, val));
}

/***********************************************************/
/* Evaluation function: this takes a user defined function. */
/* Each time this is changed, the code has to be recompiled. */
/* The current function is: x[1]^2-x[1]*x[2]+x[3]          */
/***********************************************************/

void evaluate(void)
{
  FILE *fp;
  float dayoutLAI[LAI_NUM],satLAI[LAI_NUM];
  int i, j;

  printf("evalutate\n");

  // 1) For each popsize, create dbf.epc and bak log file
  for(i=0;i<POPSIZE;i++)
  {
    printf("[****%d, %d*****]\n",generation, i);
    createDBF_file(i,generation,NVARS,population[i].gene,population[i].digit);
    // 2) run BGC model
```

```
        char strpath1[256];
        getCurrPath(strpath1);
        //strcat(strpath1, "/bgc411.exe MaeMoh_spinup.ini");
        strcat(strpath1, "/bgc411.exe MaeMoh_spinup.ini > bgclog_history/bgclog");
        createFileName(strpath1, generation, "_");
        createFileName(strpath1, i, ".txt");
        //printf("Run BGC model %d_%d: %s\n", generation, i, strpath1);
        system(strpath1);
        //printf("Run BGC model %d_%d: %s\n", generation, i, strpath1);

        // 3) convert dayout file and save into history
        //printf("***START FILE CONVERSION***\n");
        convertDayout(i, generation);
        // read dayout file and store into dayoutLAI array
        //printf("***START READING DAYOUT FILE***\n");
        readDayout(LAI_PERIOD, LAI_NUM, DAYSINYEAR, dayoutLAI);
        //testFloat(LAI_NUM, dayoutLAI);
        // read satLAI file and store into satLAI array
        //printf("***START READING SATLAI FILE***\n");
        readSatLAI(LAI_NUM, satLAI);
        //testFloat(LAI_NUM, satLAI);

                        for(j=0; j<NVARS; j++) printf("*%lf\n", population[i].gene[j]);
                        // 4) calculate fitness using dayoutLAI and satLAI
        //printf("***START CALCULATING FITNESS***\n");
        population[i].fitness = getFitness(LAI_NUM, DAY_ERR_NUM, day_err, dayoutLAI, satLAI);
        printf("fitness %0.5lf\n", population[i].fitness);
    }//end for
}

/**************************************************************/
/* Keep_the_best function: This function keeps track of the   */
/* best member of the population. Note that the last entry in */
/* the array Population holds a copy of the best individual.  */
/**************************************************************/

void keep_the_best()
{
            int mem;
            int i;
            cur_best=0; /* stores the index of the best individual */
            //population[POPSIZE].fitness = population[0].fitness;
            for (mem=0;mem<POPSIZE;mem++)
            {
                        if (population[mem].fitness > population[POPSIZE].fitness)
                        {
                                    cur_best=mem;
                                    population[POPSIZE].fitness=population[mem].fitness;
                        }
            }

            //printf("%d Best POP:\t%d\n", generation, cur_best);
/* once the best member of the population is found, copy the genes */
            for (i=0;i<NVARS;i++)
            {
                        population[POPSIZE].gene[i]=population[cur_best].gene[i];
                        //printf("\tBest gene:\t%lf\n", population[cur_best].gene[i]);
            }//end for
}
```

```
/**************************************************************/
/* Elitist function: The best member of the previous generation */
/* is stored as the last in the array. If the best member of    */
/* the current generation is worse then the best member of the  */
/* previous generation, the latter one would replace the worst  */
/* member in the new population.                                */
/**************************************************************/

void elitist()
{
        int i;
        double best, worst;          /* best and worst fitness values */
        int best_mem, worst_mem; /* indexes of the best and worst member */

best=population[0].fitness;
worst=population[0].fitness;
for(i=0;i<POPSIZE-1;i++)
{
        if(population[i].fitness>population[i+1].fitness)
        {
                if(population[i].fitness>=best)
                {
                        best=population[i].fitness;
                        best_mem=i;
                }
                if(population[i+1].fitness<=worst)
                {
                        worst=population[i+1].fitness;
                        worst_mem=i+1;
                }
        }
        else
        {
                if(population[i].fitness<=worst)
                {
                        worst=population[i].fitness;
                        worst_mem=i;
                }
                if(population[i+1].fitness>=best)
                {
                        best=population[i+1].fitness;
                        best_mem=i+1;
                }
        }
}

/* If best individual from the new populationis better than */
/* the best individual from the previous population, then   */
/* copy the best from the new population; else replace the  */
/* worst individual from the current population with the    */
/* best one from the previous generation                     */

if (best>=population[POPSIZE].fitness)
{
        for(i=0;i<NVARS;i++)
                population[POPSIZE].gene[i]=population[best_mem].gene[i];
        population[POPSIZE].fitness=population[best_mem].fitness;
}
else
{
```

```
		for(i=0;i<NVARS;i++)
			population[worst_mem].gene[i]=population[POPSIZE].gene[i];
		population[worst_mem].fitness=population[POPSIZE].fitness;
}
}

/***********************************************************/
/* Selection function: standard proportional selection for   */
/* maximization problems incorporating elitist model - makes */
/* sure that the best member survives.                 */
/***********************************************************/

void selecta(void)
{
		int mem, i, j, k;
		double sum=0.;
		double p;

/* find total fitness of the population */
		for (mem=0;mem<POPSIZE;mem++)
		{
			sum+=population[mem].fitness;
		}

/* calculate relative fitness */
		for(mem=0;mem<POPSIZE;mem++)
		{
			population[mem].rfitness=population[mem].fitness/sum;
		}
		population[0].cfitness=population[0].rfitness;

/* Calculate cumulative fitness */
		for(mem=1;mem<POPSIZE;mem++)
		{
			population[mem].cfitness=population[mem-1].cfitness+population[mem].rfitness;
		}

/* finally select survivors using cumulative fitness. */
		for(i=0;i<POPSIZE;i++)
		{
			srand((unsigned)time(NULL)+seed);
			seed = (seed+1)%1000;
			//printf("seed: %d\n", seed);
			p=rand()%1000/1000.0;
			//printf("Selecta P: %lf | %lf\n", p, population[0].cfitness);
			if(p>population[0].cfitness)
				newpopulation[i]=population[0];
			else
			{
				for(j=0;j<POPSIZE;j++)
				{
					//printf("\t%lf | %lf\n", p, population[j].cfitness);
					if(p<=population[j].cfitness && p>population[j+1].cfitness)
					{
						//printf("\tSelected %d %lf >= %lf\n", j+1, p,
population[j].cfitness);
						//printf("\t\t&& %lf < %lf\n", p, population[j+1].cfitness);
						newpopulation[i]=population[j+1];
					}//end if
				}//end for
```

```
                }
        }

/* Once a new population is created, copy it back */

        for(i=0;i<POPSIZE;i++)
                population[i]=newpopulation[i];

        //printf("\tSelecta POP\n");
        //for(i=0;i<POPSIZE;i++)
                //for(j=0; j<NVARS;j++) printf("\t\t[%d,%d] %lf\n", i, j, population[i].gene[j]);
}

/***************************************************************/
/* Crossover selection: selects two parents that take part in */
/* the crossover. Implements a single point crossover.       */
/***************************************************************/

void crossover(void)
{
        int i, mem, one;
        int first=0; /* Count of the numbers of members chosen */
        double x;

        for(mem=0;mem<POPSIZE;mem++)
        {
                srand((unsigned)time(NULL)+seed);
                seed = (seed+1)%1000;
                //printf("seed: %d\n", seed);
                x=rand()%1000/1000.0;
                //printf("Crossover x: %lf | %lf\n", x, PXOVER);
                if(x<PXOVER)
                {
                        ++first;
                        if(first%2==0)
                        {
                                Xover(one,mem);
                                //printf("\t\tXOver: %d,%d\n", one, mem);
                        }
                        else
                                one=mem;
                }
        }
}

/***************************************************************/
/* Crossover: performs crossover of the two selected parents. */
/***************************************************************/

void Xover(int one, int two)
{
        int i;
        int point; /* crossover point */

        /* select crossover point */
        if(NVARS>1)
        {
                if(NVARS==2)
                        point=1;
                else{
```

```
				srand((unsigned)time(NULL)+seed);
				seed = (seed+1)%1000;
				//printf("seed: %d\n", seed);
				point=(rand()%(NVARS-1))+1;
			}
			for(i=0;i<point;i++)
				swap(&population[one].gene[i], &population[two].gene[i]);
		}
}

/**************************************************************/
/* Swap: A swap proceudre that helps in swapping two variables */
/**************************************************************/

void swap(double *x, double *y)
{
	double temp;

	temp=*x;
	*x=*y;
	*y=temp;

}

/**************************************************************/
/* Mutation: random uniform mutation. A variable selected for */
/* mutation is replaced by a random value between lower and  */
/* upper bounds of this variable.                  */
/**************************************************************/

void mutate(void)
{
	int i,j, dig;
	double lbound, hbound, x;
	for(i=0;i<POPSIZE;i++)
	{
		for(j=0;j<NVARS;j++)
		{
			srand((unsigned)time(NULL)+seed);
			seed = (seed+1)%1000;
			//printf("seed: %d\n", seed);
			x=rand()%1000/1000.0;
			//printf("Mutate x: %lf | %lf\n", x, PMUTATION);
			if(x<PMUTATION)
			{
				/* Find the bounds on the variable to be mutated */
				if(isRandom(j, RAND_NUM, arr_rand)==1)
				{
					//printf("\t\tMutate POP: %d\t%lf\n", i,
population[i].gene[RANDVAR]);
					lbound=constpopulation[0].lower[j];
					hbound=constpopulation[0].upper[j];
					dig = population[i].digit[j];
					population[i].gene[j]=randval(lbound,hbound, dig);
					//printf("\t\t\tNew Val: %lf < %lf > %lf\n\n",
constpopulation[0].lower[RANDVAR], population[i].gene[RANDVAR],
constpopulation[0].upper[RANDVAR]);
				}//end if
			}//end if
		}//end for
```

```
		}//end for
}

/**************************************************************/
/* Report function: reports progress of the simulation. Data  */
/* dumped into the output file are separated by commas        */
/**************************************************************/

void report(void)
{
	int i;
	double best_val;  /* best population fitness */
	double avg;             /* avg population fitness */
	double stddev;          /* std. deviation of population fitness */
	double sum_square;      /* sum of square for std. calc */
	double square_sum;      /* square of sum for std. calc */
	double sum;             /* total population fitness */

	sum=0.0;
	sum_square=0.0;

	for(i=0;i<POPSIZE;i++)
	{
		sum+=population[i].fitness;
		sum_square+=population[i].fitness*population[i].fitness;
	}

	avg=sum/(double)POPSIZE;
	square_sum=avg*avg*(double)POPSIZE;
	stddev=sqrt((sum_square-square_sum)/(POPSIZE-1));
	best_val=population[POPSIZE].fitness;

	fprintf(galog, "\n%5d\t%lf\t%lf\t%lf\n\n", generation, best_val, avg, stddev);
}

/**************************************************************/
/* Main function: Each generation involves selecting the best */
/* members, performing crossover & mutation and then          */
/* evaluating the resulting population, until the terminating */
/* condition is satisfied.                                    */
/**************************************************************/


int main(void)
{
	int i;
	if((galog=fopen("galog.txt","w"))==NULL)
	{
		exit(1);
	}
	generation=0;

	printf("GAreal is starting...\n");

	fprintf(galog, "\n generation best average standard \n");
	fprintf(galog, " number value fitness deviation \n");
	fprintf(galog, "\ngen\tbest_val\tavg\tstddev\n\n");

	initialize();
	evaluate();
```

```
	keep_the_best();
	printf("POP fitness %0.5lf\n", population[POPSIZE].fitness);
	while(generation<MAXGENS)
	{
		generation++;
		selecta();
		crossover();
		mutate();
		report();
		evaluate();
		elitist();
		printf("POP fitness %0.5lf\n", population[POPSIZE].fitness);
	}
	fprintf(galog,"\n\n Simulation completed\n");
	fprintf(galog,"\n Best member: \n");

	for(i=0;i<NVARS;i++)
	{
		fprintf(galog,"\n var(%d) = %3.3f",i,population[POPSIZE].gene[i]);
	}
	fprintf(galog,"\n\n Best fitness = %3.3f\n\n",population[POPSIZE].fitness);
	fclose(galog);
	printf("Simulation completed\n");

	return 0;
}

/************************************************************/
```