



DIPARTIMENTO DI SCIENZE DELL'AMBIENTE FORESTALE E DELLE SUE RISORSE

DOTTORATO DI RICERCA IN

ECOLOGIA FORESTALE – XXI CICLO

**IMPROVEMENT OF LPJ DYNAMIC GLOBAL
VEGETATION MODEL BY MEANS OF
NUMERICAL ASSIMILATION METHODS:
POSSIBLE IMPLICATIONS FOR REGIONAL
CLIMATE MODELS**

Settore Scientifico Disciplinare – AGR/05

Dottorando:
ALESSANDRO ANAV

Coordinatore:
Prof. **PAOLO DE ANGELIS**

Tutore:
Prof. **RICCARDO VALENTINI**

ABSTRACT

Some of the feedbacks between vegetation and climate have been studied in the Euro-Mediterranean area using both models and data.

The Lund-Potsdam-Jena (LPJ) Dynamic Global Vegetation Model describes the water, carbon, and energy exchange between land surface and atmosphere by means of a given set of parameters and input variables. In order to retrieve the underlying probability density function of some key-model parameters controlling water and carbon cycle as well as to improve the efficiency of LPJ to simulate water and carbon fluxes a data assimilation system has been developed; it is based on a Bayesian approach that consistently combines prior knowledge about parameters with observations. Daily values of evapotranspiration and gross primary production, measured with eddy covariance technique in ten different *CarboeuropeIP* sites, have been compared with modeled data in order to constrain parameter values and uncertainties.

Results show how data assimilation is a useful tool to improve the ability of the model to simulate correctly water and carbon fluxes at local scale: after the inversion, in fact, LPJ successfully matches the observed seasonal cycle of the diverse fluxes, and corrects for the prior misfit to day-time GPP and ET.

The impact of land cover change on regional climate have been analyzed using the mesoscale model RegCM3. Three different simulations have been performed to assess the effects of an hypothetical deforestation and afforestation on climate.

Results show how land cover changes have a substantial impact on dynamic and thermodynamic, and how also area does not affected by land cover changes shows a significant variability in some climatic fields. Finally, the land cover changes have an important impact on the extreme events.

This thesis highlights how vegetation dynamics and climate influence each other. For such reason to improve simulation results we should develop fully coupled models that take into account some of the most important feedbacks between land surface and atmosphere.

CONTENTS

Abstract

1 INTRODUCTION

1.1	Background	1
1.2	Objectives of this thesis	2

2 LAND SURFACE-CLIMATE INTERACTIONS

2.1	Land surface processes	4
2.2	The role of terrestrial biosphere on the global carbon cycle	8
2.3	Dynamic Global Vegetation Models	14
2.4	The concept of data assimilation	19

3 A BAYESIAN INVERSION OF LPJ DYNAMIC GLOBAL VEGETATION MODEL

3.1	Introduction	22
3.2	Methods and Data: An implementation of a data assimilation system into LPJ	24
3.2.1	Optimization algorithms	24
3.2.1.1	Conjugate gradient methods	25
3.2.1.2	Metropolis Algorithm	29
3.2.2	The optimization schemes	33
3.2.3	Model description and input data	35
3.2.4	Initial model parameters values and prior uncertainties	41
3.2.5	Eddy-covariance flux data	46
3.3	Results: Changes in model parameters and fluxes after a Monte Carlo sampling	52
3.3.1	Introduction	52
3.3.2	Comparing different optimization strategies and performances	53

3.3.3	Effect of the different parameterizations on water and carbon fluxes	58
3.3.4	Interstation and interannual variability	65
3.3.5	Final-optimized parameters and uncertainty	72
3.3.6	Validation against eddy data for year 2003	76
3.3.7	Global effects of the new parameterization	79
3.4	Discussions and conclusions	87
4	EFFECTS OF LAND COVER CHANGES ON CLIMATE OVER EURO MEDITERRANEAN AREA	
4.1	Introduction	91
4.2	Model, data, and experimental setup	93
4.2.1	Model description	93
4.2.2	Potential vegetation	94
4.3	Results: Impacts of land cover change on climate	97
4.3.1	Comparison of domain-averaged climate	97
4.3.2	Spatial differences in seasonally surface climate	98
4.3.3	Differences in atmospheric circulation	104
4.3.4	Impact on extreme events	105
4.4	Discussions and conclusions	107
	Appendix	110
	Bibliography	114

I

INTRODUCTION

1.1 BACKGROUND

Over the last decades both model and observational studies have shown that the climate system is sensitive to the processes that characterize the earth's continental surface and that an accurate representation of these processes in climate models is of great importance [Alessandri *et al.*, 2007]. However, even if the interaction between atmosphere and land surface systems is an essential aspect for climate studies many uncertainties are still remaining, due to an inadequate understanding of all the processes and complex interactions involved.

The models of the coupled climate-carbon cycle system, on one hand, vary widely in predictions of future CO₂ sinks [Friedlingstein *et al.*, 2006]. The strength of this sink, which consists uptake through terrestrial ecosystems and the oceans, has important implications for creating emissions targets to reduce the likelihood of dangerous anthropogenic interference in the climate system. Thus, policy decisions must be made in the face of large uncertainties. By assimilating measurement of CO₂ fluxes into process-based biogeochemical models, we can constrain some model parameters and also decrease the misfit between simulation results and observations; this step could lead to a reduction of the uncertainty in future predictions, and hence we should be able to deliver good estimates of the carbon sinks and sources.

On the other hand, an improved knowledge and description of land surface water, energy and carbon conditions plays a pivotal role for improving land ecosystem and climate prediction. For example, soil moisture is a crucial variable for climate models because it influences the partitioning of available energy into sensible and latent heat fluxes and hence the evolution of the lower atmospheric conditions.

During the last few years, soil-vegetation schemes coupled to global climate models have become a fundamental tool to improve our knowledge of these processes. Global Circulation Models (GCMs) have been used to describe soil-atmosphere interactions and feedbacks in a

wide variety of climatic conditions, including studies related to climate change. More recently, Regional Climate Models (RCMs) have also started to be used for these studies. Regional numerical models are an interesting tool to analyze surface processes highly related to regional scales, as many of the aspects related to hydrology and surface water budget mechanisms, and their transmission to the free atmosphere through the boundary layer. Thus, regional climate models have been used to study impacts of deforestation processes, changes in land-use, local and non-local changes in precipitation due to soil moisture modifications, initial soil moisture conditions influence on precipitation for long time periods (even months), relations of soil moisture and rainfall for drought or flood conditions, impacts of regional anthropogenic vegetation changes.

The results of these studies suggest that the feedbacks between land surface and atmosphere are key determinants of climate at a range of spatial (local to global) and temporal (seasonal to centennial) scales. In fact, many of the properties involved (vegetation type and cover, soil moisture, and snow cover) evolve continuously in response to atmospheric-climatic forcing, while the initial forcing may be amplified or dampened as a consequence of their interaction.

For such reason, the necessity of more realistic and accurate computations of the exchanges of energy, momentum, water and carbon between the land surface and the atmosphere has been leading to continuous developments in land surface models. A parameterization of the vegetation included in the global circulation models or in the regional climate models allow us to better simulate the evolution of surface parameters such as roughness length, albedo, and surface–soil moisture [Alessandri *et al.*, 2007]. Furthermore, the inclusion of a realistic vegetation allows a description of the function of roots, of the physiological control of transpiration, and of the water interception by the vegetation canopies, which is quickly evaporated back to the atmosphere and this can improve our simulation results.

1.2 OBJECTIVES OF THIS THESIS

The aim of this Ph.D. thesis is to study some of the feedbacks between vegetation and climate in the Euro-Mediterranean area using both models and data. More precisely, two different studies have been performed: in the first case have a data assimilation system have been developed to constrain some model parameters controlling the exchange of water and carbon fluxes between vegetation and atmosphere, while in the latter case is analyzed how theoretical land cover changes affect the climate.

To improve our knowledge on the influence of local-climate conditions on the parameterization of some model processes, we constrained twelve relevant LPJ dynamic global vegetation model parameters controlling water and carbon cycle; this allowed as well to minimize the mismatch between simulations results and measurement of water and carbon fluxes in the Euro-Mediterranean area. The improved knowledge of present-day parameter variability can be used in future to improve regional scenario simulations for the 21st century.

To achieve this objective we used multiple sites eddy data from the CarboEuropeIP network to develop a framework for nonlinear parameter estimation and to retrieve the shape of the underlying probability density function (PDF) of twelve model parameters controlling photosynthesis and evapotranspiration in the Lund-Potsdam-Jena Dynamic Global Vegetation Model (LPJ-DGVM) [Sitch *et al.*, 2003]. The results of the multi sites optimization are used to analyze the spatial variability of parameters within and between different PFTs, and also to find out possible systematic defects in the model parameterization.

The analyses of parameter distribution as a function of dominant vegetation and climate, and in comparison to prior knowledge about the parameters from leaf-level will advance our understanding of how global models can represent specific processes of different ecosystems.

Our hypotheses are:

- 1) that parameters of the Farquhar model of photosynthesis, and of water uptake and potential evapotranspiration change only slightly at ecosystem scale and between sites with the same vegetation type or climate forcing.
- 2) that the coupling between carbon and water fluxes through canopy conductance is reliable by the LPJ-DGVM.

On the other hand, to evaluate how the vegetation affect the climate, different simulation that make use of some hypothetical land cover changes have been performed by mean of a regional climate model (RegCM3) [Giorgi *et al.*, 1990; Giorgi *et al.*, 1993a, b]. In mesoscale models the land surface represents the lowest limit of the atmosphere, but the vegetation dynamics have been poorly described. For such reason here we focus on the effect of the changes in vegetation coverage on climate

The structure of the thesis is as follows. In the chapter 2 will be briefly discussed the main interactions between terrestrial ecosystems and atmosphere. In Chapter 3 will be described the construction and application of a data assimilation system, while in Chapter 4 will be evaluated the impact of theoretical land cover change on climate.

II

LAND SURFACE-CLIMATE INTERACTIONS

2.1 LAND SURFACE PROCESSES: TERRESTRIAL ECOSYSTEMS- ATMOSPHERE INTERACTIONS

Terrestrial ecosystems-atmosphere interactions refer to exchange of heat, moisture, traces gases, aerosols, and momentum between land surfaces and the overlying air [*Pielke et al*, 1998].

Terrestrial ecosystems and climate influences one another on timescale ranging from seconds to millions of years [*Sellers et al.*, 1995; *Pielke et al.*, 1998]. Ecosystems influence weather and climate over period of seconds to years through exchanges of energy, moisture, and momentum between the land surface and the atmosphere and the changes in global-scale atmospheric circulation that can result from changes in these fluxes [*Pielke et al.*, 1998]. Ecosystem structure and function is strongly determined on timescales of decades to centuries by climate influences, primarily through temperature ranges and water availability. On timescales of thousands of years, glacial-interglacial cycles probably involve coupled changes in the geographical distribution of the terrestrial ecosystems, surface albedo, biogeochemical cycling, and climate in response to changing solar forcing. On even longer geological timescales (millions of years), terrestrial ecosystems and the Earth's climate have evolved together through such mechanisms as changes in the biochemistry and the composition of the atmosphere [*Pielke et al.*, 1998].

Land covers about 30% of the surface area of the Earth; the land surface is considered as the lower level for the planetary boundary layer (PBL). The PBL represent the lowest part of the troposphere, where wind, temperature, and humidity are strongly influenced by the surface processes. These processes are so tightly intertwined that they cannot be treated separately. Net radiation is partitioned primarily among three major avenues of energy exchange between the

ecosystem and the atmosphere: ground heat flux, latent heat flux and sensible heat flux [*Chapin et al.*, 2002].

The ground heat flux represents the heat transferred from the surface into and out of the soil; it is negligible over a day in most temperate and tropical ecosystems because the heat conducted into the soil during the day is balanced by heat conducted back to the surface at night. The magnitude of ground heat flux depends on the thermal gradient between the soil surface and deep soils and the thermal conductivity of soils, which is greatest in soils that are wet and have a high bulk density. In contrast to the temperate soils, permafrost regions of the arctic and boreal forest have substantial ground heat flux, due to the strong thermal gradient between the soil surface and the permafrost [*Chapin et al.*, 2002].

Solar energy drives also the hydrological cycle through the vertical transfer of water from Earth to the atmosphere via evapotranspiration (or latent heat), the sum of evaporation from surface and transpiration, which is the water loss by plants. So, the latent heat flux is the energy transferred to the atmosphere when water is transpired by plants or evaporates from leaf or soil surfaces. This heat is transported from the surface into the atmosphere by convection. Evapotranspiration accounts for 75% of the turbulent energy transfer from the Earth to the atmosphere and is therefore a key process in Earth's energy [*Chapin et al.*, 2002].

Finally, sensible heat flux is the heat that is initially transferred to the near-surface atmosphere by conduction and to the bulk atmosphere by convection: it is controlled in part by the temperature differential between the surface and the overlying air. Air close to the surface becomes warmer and more buoyant than the air immediately above it, causing this parcel of air to rise. Mechanical turbulence is caused by winds blowing across a rough surface: it generates eddies that transport warm moist air away from the surface and bring cooler drier air from the bulk atmosphere back toward the surface. Surface turbulence is the major process that transfers latent and sensible heat between the surface and the atmosphere [*Chapin et al.*, 2002].

There are important interactions between latent and sensible heat fluxes from ecosystems. The consumption of heat by the evaporation of water cools the surface, thereby reducing the temperature differential between the surface and the air that drives sensible heat flux. Conversely, the warming of surface air by sensible heat flux increases the quantity of water vapour that the air can hold and causes convective movement of moist air away from the evaporating surfaces. Both of these processes increase the vapour pressure gradient that drives evaporation. Because of these interdependencies, surface moisture has a strong effect on the Bowen ratio—that is, the ratio of sensible to latent heat flux [*Chapin et al.*, 2002].

Bowen ratios range from less than 0.1 for tropical oceans to greater than 10 for deserts, indicating that either latent heat flux or sensible heat flux can dominate the turbulent energy transfer from

ecosystems to the atmosphere, depending on the nature of the ecosystem and the climate. In general, ecosystems with abundant moisture have higher rates of evapotranspiration and therefore lower Bowen ratios than do dry ecosystems. Similarly, ecosystems dominated by rapidly growing plants, which have high transpiration rates, have proportionately lower sensible heat fluxes and low Bowen ratios. Strong winds and/or rough canopies, which generate surface turbulence, tend to prevent a temperature build-up at the surface and therefore reduce sensible heat flux and Bowen ratio. For these reasons, energy partitioning varies substantially both seasonally and among ecosystems. The Bowen ratio determines the strength of the linkage between the energy budget and the hydrologic cycle, because it is inversely related to the proportion of net radiation that drives water loss from ecosystems: the lower the Bowen ratio, the tighter the linkage between the energy budget and the hydrologic cycle [*Chapin et al.*, 2002].

Besides influencing the atmosphere by transpiration and the associated partitioning of surface heat fluxes into latent and sensible contributions, vegetation also affects the surface albedo and the roughness length [*Heck et al.*, 2001].

The albedo over vegetated land surfaces can vary from very low values (10-15% over humid tropical forests) to somewhat larger values (15-20% over herbaceous vegetation) [*Hartmann*, 1994]. The vegetation albedo is a function of plant structural and optical properties and the leaf area index [*Dickinson*, 1983]. As the amount of leaf area increases, light absorption increases inside a canopy and the reflection from background soil decreases. This results in an overall decrease in surface albedo. A similar situation is observed in the case of snow in tree-covered landscapes. Land surface with fully covered snow has high values of surface albedo (~0.8). The high albedo is masked in the presence of trees and can reach significantly lower values (~0.2-0.4) depending on vegetation cover type.

The roughness of a canopy surface influences the partitioning of sensible and latent heat fluxes between the land surface and the overlying air. Changes in vegetation height and leaf area exert a larger drag force on the atmospheric boundary layer and this influences the atmosphere dynamically as winds blow over the land's surface. Roughness is determined by both topography and vegetation [*Hartmann*, 1994].

Besides vegetation, also soils play a pivotal role representing an important seasonal water reservoir for the hydrological cycle. In midlatitudes soil plays a similar role to that of the oceans, but instead of storing heat, it stores precipitation in winter, which moistens then the atmosphere in summer via evapotranspiration [*Heck et al.*, 2001]. The associated seasonal storage of water in the soil introduces long-term memory effects with timescales of several months which interact with the typical atmospheric timescales of few days.

The study of the feedback mechanisms of the coupled land-atmosphere system has become increasingly important in recent years [Entekhabi, 1995; Betts *et al.*, 1996], and their numerical parameterizations in atmospheric models become more and more refined [e.g., Dickinson *et al.*, 1986; Sellers *et al.* 1986]. Extensive, satellite-based data sets are currently used to describe land surface parameters and monitor their seasonal and interannual variations. Feedback mechanisms between the land and the atmosphere are also relevant to climate change issues, since they may modulate and interact with anthropogenic changes. Anthropogenic CO₂ emissions and the associated temperature changes [Houghton *et al.*, 1995] may affect the physiological characterization of plants species [Sellers *et al.*, 1996; Betts *et al.*, 1997; Bounoua *et al.*, 1999; Heck *et al.*, 2001].

General circulation models and regional climate models have increasingly been used to investigate the atmospheric and climate response to imposed global or regional changes of surface parameters (e.g., tropical deforestation, see Heck *et al.* [2001] for detailed references). Modeling of meteorological flows requires the use of conservation equations for fluid velocity, heat, mass of dry air, water substance in its three phases, and many other natural and anthropogenic atmospheric constituents. The characterization of biospheric processes in these models, however, has been limited to simple representation where most aspects of the soil and vegetation are prescribed. Stomatal conductance responds to atmospheric inputs of solar radiation, air temperature, air relative humidity, precipitation, air carbon dioxide concentration, and to soil temperature and moisture. Till few years ago, in meteorological models, these were the only meteorological variables to which the vegetation and the soil dynamically respond [Pielke *et al.*, 1998].

The last generation of GCM and RCM has a complex vegetation or its fully coupled with a complex land surface model. Current GCM and RCM, however, also lack dynamical vegetation and carbon cycling, so are unable to take account of feedbacks related to the evolution of vegetation structure, composition and growth as conditions change. The effects of regional-to-global feedbacks are highly relevant when assessing the carbon cycle and the greenhouse gasses forcing and can be a significant source of uncertainty [Morales *et al.*, 2007].

The modeling of terrestrial ecosystems involves the short-term response of vegetation and soils to atmospheric effects, and the longer-term evolution of species composition, biome dynamics, and nutrient cycling associated with landscape and soil structure changes. The assimilation of carbon resulting in the growth of vegetation, and its subsequent release during decay has been a focus of these models. The spatial scale of these simulations have ranges from patch size (microscale) to biome (mesoscale) scales.

2.2 THE ROLE OF TERRESTRIAL BIOSPHERE ON THE GLOBAL CARBON CYCLE

Carbon dioxide (CO₂) is a naturally abundant trace gas in the atmosphere. Through its radiative properties it is (besides atmospheric water vapour) the most important greenhouse gas. Because of the natural greenhouse effect the mean global surface temperature amounts to around +15 °C compared to -18 °C without any climate relevant trace gases in the atmosphere and therefore is a necessity for our life on earth.

Although many elements are essential to living matter, carbon is the key element of life on Earth. The biogeochemical cycle of carbon is necessarily very complex, since it includes all life forms on Earth as well as the inorganic carbon reservoirs and the link between them. So, to understand and project future changes in the global carbon cycle, it is necessary to understand its underlying elements and processes.

The carbon cycle consists of three main compartments (*Figure 2.1*): the atmosphere, the oceans, and the biosphere. Each compartment consists of different carbon pools and exchanges carbon at different rates.

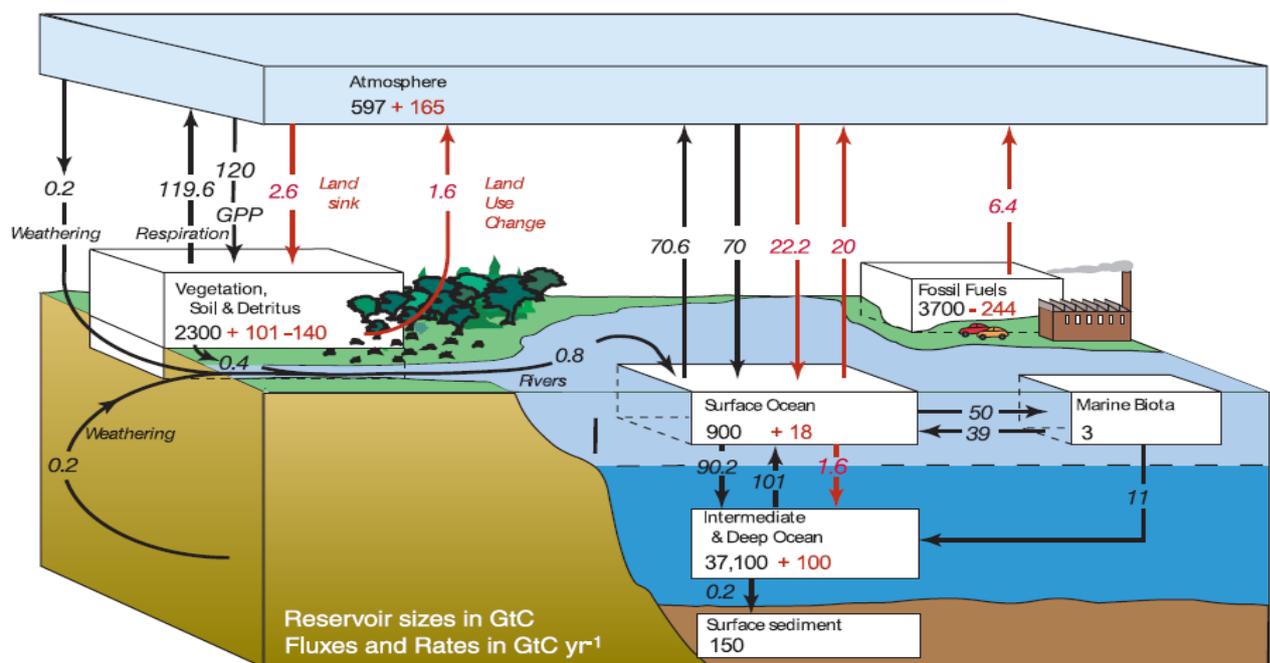


Figure 2.1. The global carbon cycle for the 1990s, showing the main annual fluxes in PgC yr⁻¹: pre-industrial 'natural' fluxes in black and 'anthropogenic' fluxes in red. Within the boxes, black numbers give the preindustrial size of the reservoirs and red numbers denote the changes resulting from human activities since preindustrial times. For the land sink, the first red number is an inferred terrestrial land sink whose origin is speculative; the second one is the decrease due to the deforestation (taken from Denman et al. [2007]).

The largest amount of carbon (C) by far (about 38,000 Pg C) [Gruber *et al.*, 2004] is stored in the middle and deep ocean (*Figure 2.1*). This C is, however, relatively inert and, as such, less relevant for the C cycle in coming decades [Bolin *et al.*, 2000; Gruber *et al.*, 2004]. Smaller but still considerably large pools are found in the terrestrial biosphere (2100–3000 Pg C), the surface layer of the oceans (600 Pg C) and the atmosphere (700–800 Pg C) [Grace, 2004; Denman *et al.*, 2007].

Numerous well-replicated measurements of the composition of air bubbles trapped in Antarctic ice showed that the atmospheric CO₂ concentration remained relatively constant at a level of about 280 parts per million (ppm) [Indermühle *et al.*, 1999] during the last 10,000 years suggesting that the carbon cycle has been in quasi-equilibrium during that time.

Since industrialization (the last ≈150 years) atmospheric CO₂ concentration increased by more than 80 ppmv to a total value of around 370 ppmv today, a magnitude which has not been exceeded during the last 420,000 years at a rate which is unique for at least the last 20,000 years [Prentice *et al.*, 2001].

Fossil fuel burning and related industrial activities, as well as terrestrial carbon losses from land-use change are the causes for the rise in atmospheric carbon concentration. Atmospheric CO₂ is, however, increasing only at about half the rate of fossil fuel emissions; the rest of the CO₂ emitted either dissolves in sea water and mixes into the deep ocean, or is taken up by terrestrial ecosystems [Prentice *et al.*, 2001]. Partitioning of the terrestrial and ocean fluxes based on simultaneous measurements of CO₂ and O₂ suggest that the terrestrial biosphere sequesters up to 30% of the fossil-fuel emissions [Ciais *et al.*, 1995; Prentice *et al.*, 2001; House *et al.*, 2003].

Because of the pivotal role that carbon plays in the climate system it is critical that we understand all the processes that regulate its cycling through vegetation and develop such mechanisms into GCM in order to make good estimates on the future changes in the global carbon cycle.

The CO₂ uptake and release of the terrestrial biosphere is determined by a number of processes that are sensitive to climate, atmospheric CO₂, moisture availability, and land use. Within the terrestrial biosphere, plants take up CO₂ by diffusion through the stomata of leaves (globally about 270 Pg C yr⁻¹) [Ciais *et al.*, 1997]. More than 50% of this CO₂ diffuses back to the atmosphere without becoming part of biogeochemical processes within plants. A basic biogeochemical process within plants is photosynthesis, where CO₂ is converted under the influence of “active radiation” (400-700 nm wave length) into carbohydrates that serve as raw material for further processes. The amount of C that is fixed through photosynthesis is called *Gross Primary Production* (GPP, *Figure 2.2*). The amount that is really taken up by plants, allocated to and incorporated in new plant tissues is defined as *Net Primary Production* (NPP,

Figure 2.2). As such NPP includes all increments in the biomass of leaves, stems, branches, roots, and reproductive organs. The remaining part of the C is lost by *autotrophic respiration* (R_a , Figure 2.2). Most of the C fixed through NPP also returns back to the atmosphere through *heterotrophic respiration* (R_h , Figure 2.2), and *disturbances*. The former process is the decomposition of soil organic matter by bacteria and fungi, which consume most of the organic material that enters the soil through dying plant material. Several soil C pools can be distinguished with different C contents, chemical composition, and different bacteria and fungi composition. As result they have often different decomposition rates. The difference between R_h and NPP is called *Net Ecosystem Productivity* (NEP) or *Net Ecosystem exchange* (NEE, Figure 2.2). Negative fluxes denote a net carbon flux from the atmosphere to the terrestrial biosphere because the assimilation process dominate on respiration terms; in such case the ecosystem is considered a sink of carbon. It represents the amount of carbon that is annually stored in the terrestrial biosphere. When also accounting for C losses due to fires, land-use change, harvest, and erosion, the total C flux is called *Net Biome Productivity* (NBP).

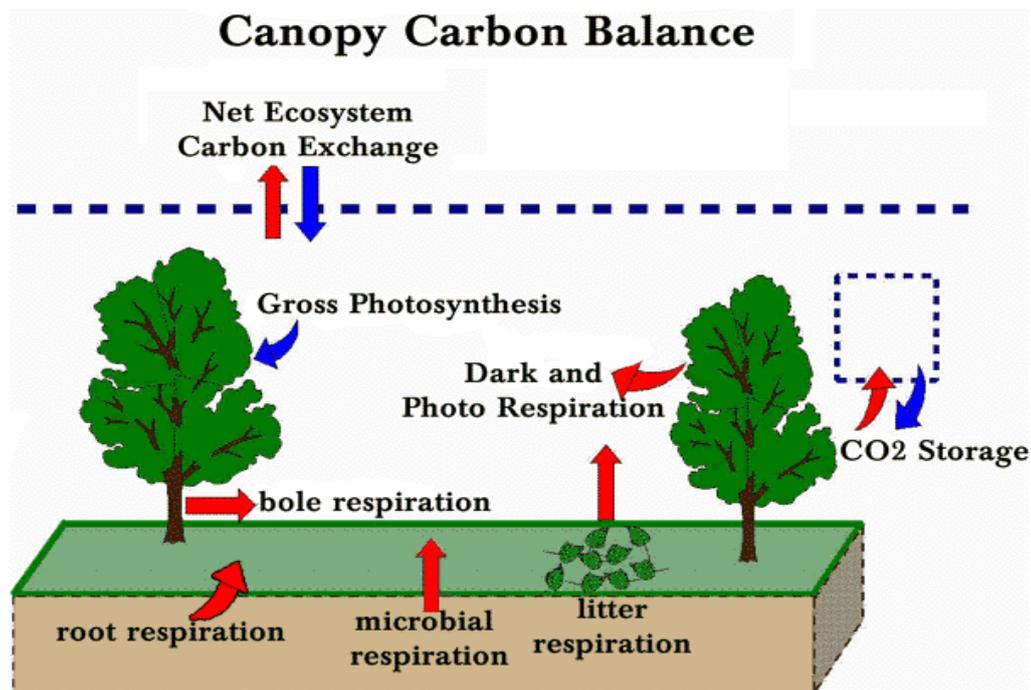


Figure 2.2. Main component of the terrestrial carbon cycle; blue arrows denote an assimilation of carbon by vegetation, while red arrows point out an emission of carbon from ecosystem to atmosphere.

Globally, the terrestrial biosphere stores about 2100–3000 Pg C, divided into 466–660 Pg C in the vegetation and 1460–2300 Pg C in soils. The total terrestrial C storage is about three times

the amount in the atmosphere and the surface layer of the ocean [Sabine *et al.*, 2004; Denman *et al.*, 2007]. The range is caused, for example, by differences in definitions (e.g. some soil compartments), the total area included and various uncertainties (especially related to the soil carbon budget, which is difficult to measure). Global annual terrestrial GPP is estimated at 120 Pg C yr⁻¹ [Denman *et al.*, 2007]. Roughly, one half of this gross primary production is lost to the atmosphere via autotrophic respiration from plant tissues [Lloyd and Farquhar, 1996; Waring *et al.*, 1998], and the remaining 53–68 Pg C yr⁻¹ Pg C are incorporated into plant tissue as terrestrial net primary production

The estimates are based on integration of field measurements, remote sensing, atmospheric measurements and modeling the historical C cycle. The range is due to uncertainties in land cover and land use [Houghton, 2003; Lambin *et al.*, 2003], and in the response of the terrestrial biosphere to environmental changes like climate, CO₂, and nitrogen fertilization.

NPP fluxes vary with the study, especially for tropical regions [Berthelot *et al.*, 2005]. Furthermore, the range is caused by the different measurement methods and the different time periods of the studies. Regarding the time period, Potter *et al.* [1999] and Nemani *et al.* [2003] showed that the global NPP increased about 6% over the last decades. The global NEP flux over the last decades is estimated at between -3 and -10 Pg C yr⁻¹ [Watson *et al.*, 2000; Cox, 2001; Cramer *et al.*, 2001; Prentice *et al.*, 2001; Grace, 2004; Schaphoff *et al.*, 2006]. Note that there is a wide range of uncertainty, especially in soil processes [Grace, 2004], and a considerable inter-annual variability [Valentini *et al.*, 2000]. Furthermore, the low end of the range often represents model results that implicitly include some effects of disturbances.

Looking at the possible future, we see that various models have projected an increasing NEP flux up to the middle of this century, followed by a stabilization [Cramer *et al.*, 2001; Scholes and Noble, 2001], a decline [Lucht *et al.*, 2006; Schaphoff *et al.*, 2006] or even a shift towards a C source [Cox *et al.*, 2004]. The decrease (and shifts towards a C source) is due the enhanced heterotrophic respiration due to the increased mean atmospheric temperature. These projections are also surrounded with substantial uncertainty due to uncertainties in future regional climate [Schaphoff *et al.*, 2006] and the response of the biosphere to future climate and atmospheric CO₂ [Cramer *et al.*, 2001; Friedlingstein *et al.*, 2006].

When CO₂ emissions due to land-use changes are excluded, the global residual terrestrial C sink is estimated to be in the range of -0.9 to -2.4 Pg C yr⁻¹ over the 1980s and -2 to -3 Pg C yr⁻¹ in the 1990s [Denman *et al.*, 2007]. Recent observations indicate that the global sink is still increasing, possibly up to -2.9 Pg C yr⁻¹ [Körner, 2003; Grace, 2004].

The C pools and fluxes are not homogeneously distributed across the world, but differ geographically, seasonally and between ecosystem types (*Table 2.1*). Ecosystem types that store large quantities of C are forests, grasslands, and wetlands.

Ecosystem	Area (10^9 ha)	Total C pool (Pg C)	NPP (Pg C yr ⁻¹)
Forests:	4.2 (28%)	1146 – 1640 (46 – 59%)	25 – 33 (41–52%)
Tropical	1.9	428 – 1032	15.2 – 21.9
Temperate	1.0	159 – 401	5.5 – 8.1
Boreal	1.3	207 – 559	2.4 – 4.2
Grasslands & savannas	3.5 (24%)	508 – 634 (21 – 25%)	14 – 20.5 (23–33%)
Agriculture	1.6 (11%)	131 – 169 (5 – 8%)	4 (7%)
Arctic tundra	1 (7%)	117 – 146 (5%)	0.5
Wetlands	0.4 (3%)	230 – 450 (8 – 13%)	0.2
Others (e.g. deserts)	4.2 (27%)	200 – 359 (8 – 12%)	4.9 (7%)
TOTAL	15	2137 – 2996	55 – 63

Table 2.1. Global C pools and NPP fluxes differentiated over ecosystems (sources: Silver [1998]; Gitay et al. [2001]; Nemani et al. [2003]; Grace [2004]; Sabine et al. [2004]; Fischlin et al. [2007]).

On the global scale, forests cover more than 4 billion hectares or about 28% of the terrestrial biosphere [FAO, 2001; Grace, 2004]. About half the forest area is located in developed countries (mostly temperate and boreal types of forests) and half in developing (mostly tropical) countries. Forest ecosystems play a pivotal role in the global C cycle because they store nearly half the terrestrial C (*Table 2.1*). If considering only the vegetation C pools, forests even store 80–90% of the carbon [Körner, 2003]. The largest fraction of this pool (i.e. about 60%) is stored in tropical forests [Sabine et al., 2004]. Note that the C pools and fluxes vary even considerably between tropical forest types, depending on the type of forest and degree of disturbance. The living biomass of tropical rain forests, for example, ranges between 160–190 Pg C ha⁻¹ compared to dry forests with only 30–60 Pg C ha⁻¹. Boreal forests also contain a considerable amount of carbon in the soil (i.e. 200–500 Pg C) [Sabine et al., 2004].

Secondly, forest ecosystems are important for the global C cycle because of the considerable C exchange between forests and the atmosphere. Almost all the forests around the world currently sequester C and, in particular, tropical forests fix into biomass up to 1200 gC m⁻² yr⁻¹. Including the C losses due to deforestation, tropical forests still represent either a small C sink [Houghton, 2002; Cramer et al., 2004; Grace, 2004].

Natural grasslands are also widely spread across the world. The total grassland area is about 3.5 billion hectares, of which 65% is located in warm and tropical regions [*Sabine et al.*, 2004]. Much grassland around the world has been converted into agricultural land over the past decades, resulting in a declining amount [*Meyer and Turner*, 1998]. Natural grasslands are important for the global C cycle because of the large extent and the considerable carbon storage, on the one hand, and their sensitivity to climate change and direct human influence on the other [*Parton et al.*, 1994]. Regarding the former, the global C storage in the living biomass of natural grasslands is 33–85 Pg C, while the total C pool in grassland soils is in the range of 279 and 559 Pg C [*Sabine et al.*, 2004]. Large differences are found across the world for grasslands too. The C storage in the living biomass of tropical grasslands is generally higher than in grasslands in temperate regions, whereas the soil C pools are comparable in both parts of the world [*Sabine et al.*, 2004]. Global NPP estimates of grasslands vary between 8.6 and 15 Pg C yr⁻¹. The productivity decreases due to human influences, causing a reduced C storage and soil erosion [*Burke et al.*, 1991; *Ojima et al.*, 1993]. The observed NPP range is, in particular, determined by the seasonal distribution of precipitation [*Ojima et al.*, 1993]. This is because many natural grasslands in the world are water limited [*Meyer and Turner*, 1998]. Because of the dependency on water, climate change (especially seasonality and precipitation) may have a considerable effect on the total C balance of grasslands.

Wetlands also store large amounts of carbon, mainly in soils (240–455 Pg C) [*Gorham*, 1991; *Sabine et al.*, 2004]. The majority of the wetlands and C storage can be found in boreal and arctic regions. Tropical wetlands are less relevant (total C storage about 70 Pg C) [*Diemont*, 1994], although the largest C densities have been observed here. Furthermore, wetlands are important for the C cycle because of the risk of a significant loss of their soil C pool under climate change. The optimum annual average temperature for C sequestration of most wetlands is between 4 and 10 °C. This can be currently observed in much of the southern-boreal and northern-temperate zones. With projected temperature increases, conditions are likely to exceed the optimum range. Numerous arctic wetlands may even disappear entirely under temperate increases in the range of 2–3 °C [*Hitz and Smith*, 2004]. Likewise, changes in precipitation affect the species composition of wetlands and, as such, the functioning [*Keddy*, 2000]. All these changes can result in a change in wetlands from a C sink into a C source.

Understanding the complexity of the terrestrial C dynamics in relation to anthropogenic activities and environmental changes, including future trends and assessment of possible policies, can be achieved through modeling. Various types of terrestrial C-cycle models have been developed, ranging from highly aggregated and simple, to complex process-based. Next section will give a brief description of the existing models.

2.3 DYNAMIC GLOBAL VEGETATION MODELS

To formulate a model is to put together pieces of knowledge about a particular system into a consistent pattern that can form the basis for (1) interpretation of the past history of the system and (2) prediction of the future of the system. In other words, simulation models use mathematical expressions to describe the behaviour of a system in an abstract manner [Bratley *et al.*, 1987]. The mathematical expressions are based on scientific theories and assumptions. Compared to the real world, the structure and processes are simplified in any simulation model. Simulation models have also other diverse purposes from predictions/projections (i.e. assessing responses to changing driving forces) described above such as education and an improved understanding/discovery of relationships between the elements of a system [Goudriaan *et al.*, 1999]. Furthermore, a general purpose of any model is to enable its users to draw conclusions about the real system by studying and analyzing the model results.

The complexity of ecological controls over all the processes that influence ecosystem carbon balance makes long-term projections of terrestrial carbon storage a daunting task. Making these projections is, however, critical to improving our understanding of the relative role of terrestrial ecosystems in the global carbon balance. Experiments that test the multiple combinations of environmental conditions influencing terrestrial carbon storage are difficult to design. Modeling allows a limited amount of empirical information to be greatly extended through simulation of complex combinations of environmental-biotic interactions [Chapin *et al.*, 2002]. One important use of ecosystem models has been to identify the key controls that govern rate estimates of the area of each patch type. Satellite imagery now provides improved estimates of the aerial extent of many patch types, but spatial and temporal variation in processes makes it difficult to find good representative sites from which data can be extrapolated [Chapin *et al.*, 2002]. This extrapolation approach can be combined with empirical regression relationships (rather than a single representative value) to estimate process rates for each patch type. Carbon pools in forests, for example, might be estimated as a function of temperature or normalized difference vegetation index (NDVI) rather than assuming that a single value could represent the carbon stocks of all forests.

Process-based models make up another approach to estimating fluxes or pools over large areas. These estimates are based on maps of input variables for an area (e.g., maps of climate, elevation, soils, and satellite-based indices of leaf area) and a model that relates input variables to the ecosystem property that is simulated by the model.

Many of key processes regulating NEP involve changes that occur over decades to centuries (*Figure 2.3*). The temporal resolution of the models must therefore be coarse, with time steps (the shortest unit of time simulated by the model) of a day, month, or year. Use of relatively long time steps such as weeks or months reduces the level of detail that can be considered.

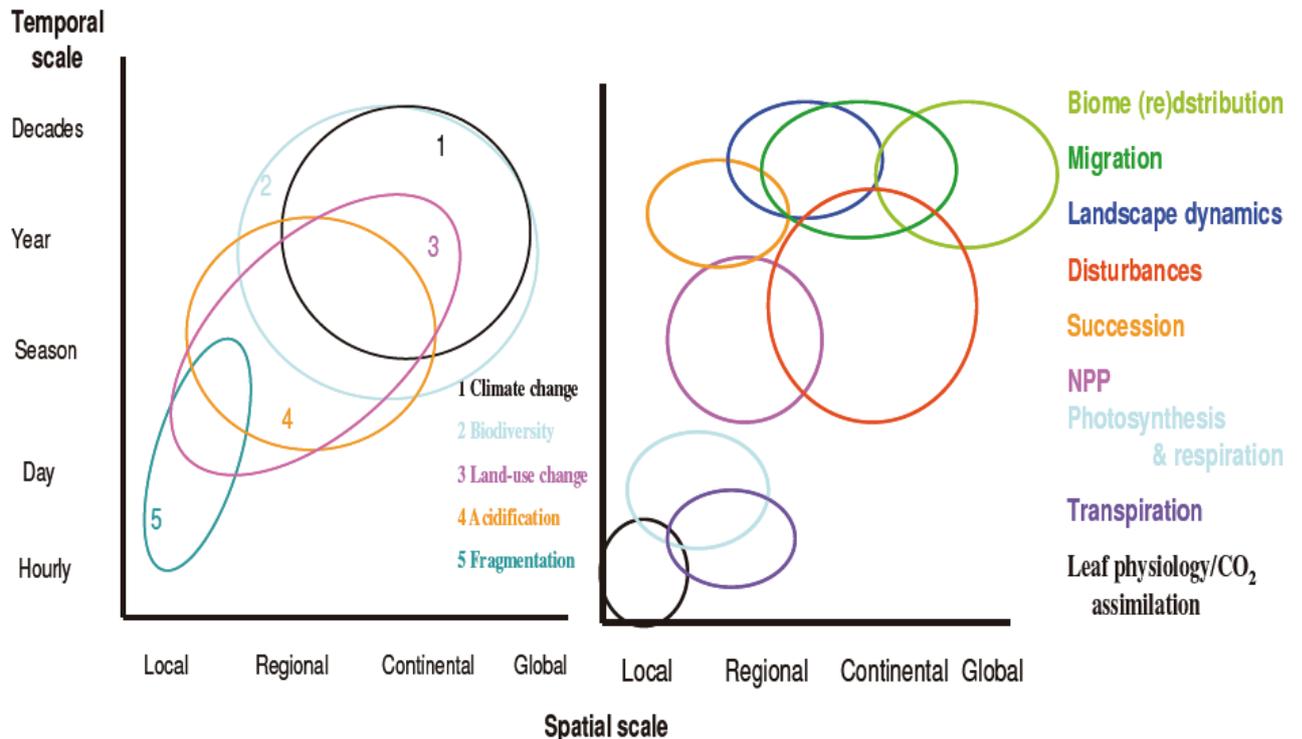


Figure 2.3. Spatial and temporal resolution of environmental processes (left panel) and carbon cycle dynamics (right panel).

The basic structure of a model of NEP must include the pools of carbon in the soils and vegetation. It must also include the fluxes of carbon from the atmosphere to plants (GPP or NPP), from plants to the atmosphere (plant respiration, harvest, and combustion), from plants to soil (litter fall), and from soil to the atmosphere (decomposition and disturbance). Models differ in the detail with which these and other pools and fluxes are represented. Plants, for example, might be considered a single pool or might be separated into different plant parts (leaves, stems, and roots), functional types of plants (e.g., trees and grasses in a savanna), or chemical fractions such as cell wall and cell contents.

So far, a broad range of models exist, and they are used to investigate the magnitude and geographical distribution of a wide numbers of variables related to the vegetation [*Cramer et al.*, 1999]. These models range in complexity from regressions between climatic variables and one or more estimates. Different models use different simplifying assumptions, and they often use

different environmental variables leading to different estimates. So, there is no single "best" model, but each model has a unique set of objectives, and the model structure must be designed to meet these objectives.

Perhaps the biggest challenge in model development is deciding which processes to include. One approach is to use a hierarchical series of models to address different questions at different scales. Models of leaf-level photosynthesis and of microclimate within a canopy have been developed and extensively tested for agricultural crops, based on the basic principles of leaf biochemistry and the physics of radiation transfer within canopies. One output of these models is a regression relationship between environment at the top of the canopy and net photosynthesis by the canopy. This environment-photosynthesis regression relationship can then be incorporated into models operating at larger temporal and spatial scales to simulate NPP, without explicitly including all the details of biochemistry and radiation transfer. This hierarchical approach to modeling provides an opportunity to validate the model output (i.e., compare the model predictions with data obtained from field observations or experimental manipulations) at several scales of temporal and spatial resolution, providing confidence that the model captures the important underlying processes at each level of resolution.

Global vegetation models (GVM) have in the past decade evolved from largely statistical correlation to more process-based, rendering greater confidence in their abilities to address questions of global change. There are generally two classes of GVMs, biogeography models and biogeochemistry models [*Haxeltine and Prentice, 1996*]: biogeochemistry models have traditionally been used to assess biogeochemical fluxes through ecosystem given prescribed distribution of ecosystem types, while biogeography models have been used to simulate the response of biome boundaries to projected changes in climate and atmospheric CO₂ [*Pan et al., 2002*]. In other words, the biogeography models place emphasis on determination of what can live where, but either do not calculate or only partially calculate the cycling of carbon and nutrients within ecosystems. The biogeochemistry models simulate the carbon and nutrient cycles within ecosystems, but lack the ability to determine what kind of vegetation could live at a given location under specific environmental conditions.

As biogeochemistry model, the Terrestrial Ecosystem Models (TEMs) were designed to simulate the carbon budget of ecosystems for all locations on Earth at 0.5° longitude by 0.5° latitude resolution (60000 grid cells) for time periods of a century or more [*McGuire et al., 2001*]. TEM has a relatively simple structure and a monthly time step, so it can run efficiently in large numbers of grid cells for long periods of time. Soil, for example, consists of a single carbon pool. The model assumes simple universal relationships between environment and ecosystem processes based on general principles that have been established in ecosystem studies. The

model assumes, for example, that decomposition rate of the soil carbon pool depends on the size of this pool and is influenced by the temperature, moisture, and C:N ratio of the soil. TEM incorporates feedbacks that constrain the possible model outcomes. The nitrogen released by decomposition, for example, determines the nitrogen available for NPP which in turn governs carbon inputs to the soil and therefore the pool of soil carbon available for decomposition. This simplified representation of ecosystem carbon dynamics is sufficient to capture global patterns of carbon storage [McGuire *et al.*, 2001], making the model useful in simulating regional and global patterns of soil carbon storage. The basic structure of a TEM is shown in *Figure 2.4*.

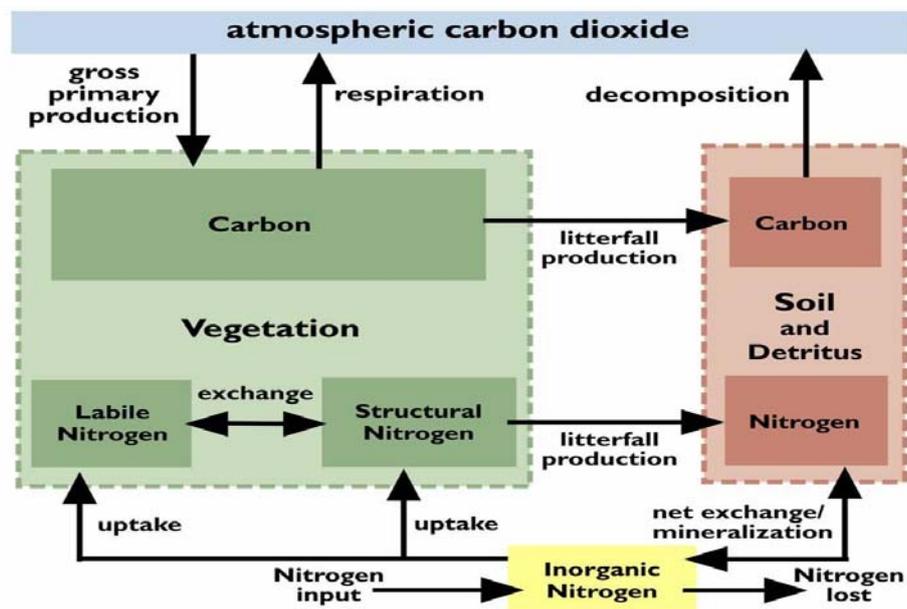


Figure 2.4. Driven from monthly climatology, TEM simulates the terrestrial carbon cycle.

To develop regional and global estimates of NPP and carbon storage, TEM normally uses geographically-referenced data sets organized at a spatial resolution of 0.5x0.5 degree to capture the spatial variability of environmental conditions across the region. Because a 0.5 degree grid cell covers a large area and environmental conditions are considered to be constant within the grid cell, NPP estimates might be improved using data sets with a finer spatial resolution.

As a matter of fact, even in a constant climate, ecosystem structure is generally unstable. Natural mortality of plants or of plant parts such as leaves and branches constantly adds biomass to the litter pool upon which other organisms thrive and thereby influence the flux of carbon from organic substances back to the atmosphere. This mortality occurs erratically, usually as a result of disturbances, and thereby creates new opportunities for other plants. Later stages of the process, such as the decomposition of dead plant matter, last much longer and may involve large

quantities of carbon that are stored for long time periods (in the case of wetlands for many millennia). For studies of the global carbon cycle, it is necessary to make at least rough estimates of these overall fluxes and their sensitivities to the environment, such as temperature and moisture balance.

Increasingly, vegetation structure has been incorporated into models as a dynamic feature of the biosphere (rather than an input), driven by climate and carbon and water fluxes: fully dynamic versions of the spatially explicit GVMs are being developed and incorporate both biogeography and biogeochemistry processes. So, in order to capture the responses of land ecosystems to climate change, processes such as resource competition, growth, mortality, and establishment have been included in terrestrial models and born a new class of model, called Dynamic Global Vegetation Model (DGVM).

These models represent vegetation not as biomes (e.g., savanna) but rather as patches of Plant Functional Types (PFT, e.g., grasses, trees). This is because many of the leaf physiological and plant allocation parameters used in ecological models cannot be measured for biomes but can be measured for individual plant types [Bonan *et al.*, 2002]. Plant functional types reduce the complexity of species diversity in ecological function to a few key plant types and provide a critical link to ecosystem processes and vegetation dynamics [Woodward and Cramer, 1996; Smith *et al.*, 1997].

Crucial for the development of DGVMs is the definition of an appropriate set of PFTs. PFTs must be few in number so that they can be parameterized in the global model, but sufficiently complex to cover at least a part of the variety in functional behavior among plants around the globe. For each PFT we need to parameterize the physiological processes. Some PFTs are defined directly on the basis of physiology, e.g. C₃ and C₄ plants, which by their different photosynthetic pathways respond in different ways to climate change. Other PFT distinctions are made according to leaf longevity, such as between deciduous and evergreen plants.

Dynamic Global Vegetation Models (DGVM), using a given number of PFT, describes the carbon and water exchange between land surface and atmosphere by means of a given set of parameters and input variables. They uses spatially referenced information on climate, elevation, soils and vegetation as well as soil and vegetation specific parameters to make estimates of water and carbon fluxes and pool sizes of terrestrial ecosystems.

A comparison of model results with field data for the location where water and carbon fluxes has been measured provides one reality check. At these sites, measurements of NEP over several years spanning a range of weather conditions provides a measure of how that ecosystem responds to variation in climate. This allows a test of the model's ability to capture the effects of ecosystem structure and climate on NEP.

2.4 THE CONCEPT OF DATA ASSIMILATION

Numerical models offer a description of the physical and ecological dynamics of the system of interest at the required time and space scales. However, models are and will always be only a simplified representation of reality. They include a number of approximations, assumptions and uncertainties that cause the model to drift away from the real state of the system that it is designed to predict. Moreover, and this is principally the case with ecological models, with the increasing complexity of the models, also their calibration becomes more and more difficult, due to the increasing number of system parameters. Finally, one of the fundamental insights in physics of the last century was that nature is not deterministic, i.e. even if the current state of a natural system would be fully known, its future state is not uniquely defined. Therefore approaches are needed, that take probabilistic considerations into account, e.g. instead of determining a theoretical future state, give rather an estimation of the most likely realisation of future events, along with its probability and an estimation of the eventual error [Evensen, 2006].

One long term interest of biogeochemical modellers is to first check relatively simple models for consistency and thereby discover the most important processes that need to be considered for large-scale predictions of biogeochemical fluxes. If these most prevailing processes are recovered and correctly parameterised, such a basic model could then be gradually modified and be extended for local process studies as well.

For ecosystem modeling it is inevitable to validate the applied model equations and to justify the associated complexity. Many terrestrial ecosystem models consist of parameterisations which are believed to describe certain biogeochemical processes within the land surface. Only few parameterisations are based upon measurements, conducted in laboratories or only in few different biomes. This is in strong contrast, for example, to physical models of the atmosphere and ocean circulation, which are mainly build on the sound theory of the Navier-Stokes equation. As a consequence, it becomes necessary to study the discrepancies between model results and observations with great care. Such investigations can be subject to three major questions:

- 1) Are the model's equations and their spatial discretization appropriate?
- 2) Are the model's parameter values optimally chosen?
- 3) Are the model's derived variables comparable with *in situ* observations?

In practice all the questions above cannot be handled separately, making a systematic approach more troublesome. For example, it is difficult to distinguish between the errors which are related

with the model equations and those which are due to the improper choice of parameter values. If one seeks for the most appropriate prognostic equations or model resolution one must answer, or at least discuss, the second and third question in advance.

With respect to the second question, data assimilation techniques are generally used to find an optimal combination of parameter values that minimise a function which describes the misfits between observation and model results, mostly named objective or cost function (J , Figure 2.5). In other words, data assimilation is a methodology which can optimise the extraction of reliable information from observations and combine it with numerical models.

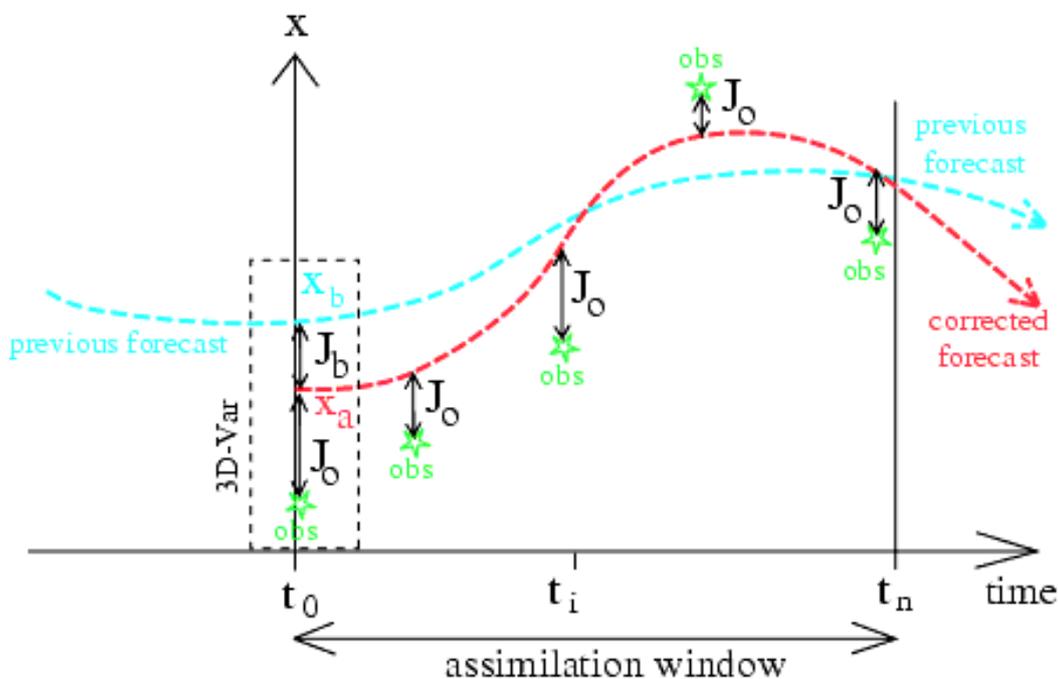


Figure 2.5. Example of data assimilation in a numerical forecasting system. Considering the observations as a long time sequence of carbon fluxes measured in a given station, we can validate a model output (previous forecast curve) with these measurements. If the model drift away from observations we can “correct” the simulations by data assimilation techniques; in other word we constrain some model parameters in order to minimize the difference between the model simulation (corrected forecast) and observations.

Thanks to the achievements in research and technology, measurements have become available for a lot of areas of the world and quite efficient and more and more accurate. Considerable effort has been spent to extend and refine the networks of data acquisition and to collect the measurement (e. g. Euroflux, Fluxnet). In particular the use of eddy covariance data has largely increased the availability of essential observational data for TEM, such as water, carbon fluxes and heat fluxes.

From a methodological point of view, we should use observed carbon and water fluxes at ecosystem scale by the eddy covariance technique to constrain the parameters derived at fine scale in the laboratory or field. This would lead to a parameterization of a global TEM which is adequate for the scale of interest. However, the analysis of parameters and their uncertainties stratified for vegetation type or climate indices allows advanced ecological interpretations, too.

The third question deals with the interpretation of biogeochemical measurements and the associated terminology. For the optimisation one usually presumes that the model counterparts to the observations are correctly calculated. In some cases it is still discussed whether a certain measurement, or its procedure, represents a specifically defined biogeochemical process sufficiently well or not. In contrast, the model's counterparts to the observations are mathematically clearly defined. Consequently, the inconsistency between a model output and a measurement result must not automatically imply that the model equations are wrong but could be subject to the inadequacy of comparing the model's diagnostic result with the measurement.

III

A BAYESIAN INVERSION OF LPJ DYNAMIC GLOBAL VEGETATION MODEL

3.1 INTRODUCTION

The terrestrial biosphere, through biophysical interaction and biogeochemical exchange with atmosphere and ocean, is able to modify local and global climate [Foley *et al.*, 1994; Prentice *et al.*, 2000]. For this purpose, terrestrial ecosystem models have been used extensively to study the processes leading to either carbon loss or gain by the land ecosystem [McGuire *et al.*, 2001; Prentice *et al.*, 2001].

One important question that models of the carbon cycle could answer is what the future atmospheric concentrations of CO₂ may be. The estimates of the terrestrial carbon budget are however highly uncertain [Friedlingstein *et al.*, 2006]. Uncertainties in the model simulations originate both from missing or non-adequate process representation, and from uncertainties in the parameters used. In fact, to run these models, a large number of parameters must be specified. To assign values to the biological parameters is especially difficult, because, unlike many physical or chemical parameters, they cannot be directly measured or regarded as constants [Matear, 1995].

So far, parameters for description of processes like photosynthesis or respiration have been estimated in laboratories or at spatial scales of centimetres to meters in the field [Farquhar *et al.*, 1980; Lloyd and Taylor, 1994]. In global models like TEMs, however, the same equations and most often also the same parameters have been applied to grid cells of 0.5° width and more,

which comprise even more than one ecosystem. The reliability of such technique is highly questionable especially because equations are most often derived empirically.

From a methodological point of view, we should use observed carbon and water fluxes at ecosystem scale by the eddy covariance technique to constrain the parameters derived at fine scale in the laboratory or field. In fact, the existence of both a model able to predict the time evolution of one or more variables and of systematic observations leads to an inverse problem; in other words, we can compute few model parameter values in an optimal way, so that the predictions of the model at the stations of measurement in the whole period of observation differ as little as possible from the observations collected.

Recently, the remarkable growing amount of eddy covariance measurements available have allowed to try to quantify uncertainty ranges for some ecosystem model parameters [Knorr, 2000; White *et al.*, 2000; Wang *et al.*, 2001; Reichstein *et al.*, 2003; Braswell *et al.*, 2005; Knorr and Kattge, 2005; Sacks *et al.*, 2006; Santaren *et al.*, 2007; Wang *et al.*, 2001]. However, eddy covariance observations have been underutilized in data assimilation studies; only few multi-site assimilation using eddy covariance data has been attempted [e.g. Wang *et al.*, 2007] and for our knowledge no published papers describe the spatial and temporal parameters variability between different ecosystems. Assimilating observations from different eddy covariance sites will reveal how coherent these parameters are both between and within PFTs. Such an analysis will also reveal whether key processes are missing or misrepresented by examining the models' ability to reproduce observed water and carbon fluxes across a range of spatial and temporal timescales.

Various techniques are used for the data assimilation, such as conjugate gradient methods, Kalman Filtering, optimal interpolation (OI), successive corrections and so on [Press *et al.*, 1992]. However, most of these techniques are designed for linear dynamics, and may fail to work correctly in the nonlinear case [Gauthier, 1992; Miller *et al.*, 1994]. Moreover, since our data have errors and the truth is imperfectly known, the problem should be considered as a probabilistic one. That is, the correct solution is not a unique set of parameters but rather a set of parameter estimates along with confidence intervals.

Standard variational methods do not naturally produce this probabilistic information; instead, they merely generate the optimum parameter values. They also require significant effort to implement. In contrast, the Monte Carlo Markov Chain (MCMC) method, based on the Metropolis-Hastings algorithm [Metropolis *et al.*, 1953], is simple to implement, very flexible (including the ability to handle nonlinear models) and produces an ensemble of parameter estimates from which statistical properties can readily be derived and probabilistic forecasts made. The substantial handicap of this method is the massive computational demand which it imposes [Hargreaves and Annan, 2002].

3.2 METHODS AND DATA: AN IMPLEMENTATION OF A DATA ASSIMILATION SYSTEM INTO LPJ

3.2.1 Optimization algorithms

In the present thesis, we developed two different algorithms into LPJ to reduce the mismatch between observations and simulations; each algorithm works minimizing a given cost function (J). The first one (*conjugate gradient methods*) is a very efficient and fast optimization algorithm that approaches the minimum often within a few iterations. However, when the cost function is a complex surface this algorithm has a tendency to terminate in a local minimum: the failures are attributed to the complex shape of the cost function, caused by non-linearities in the model which lead to discontinuities in the cost function [Matear, 1995]. For such reason there exists an uncertainty in using suitable starting conditions that make it difficult to judge if the global minimum was reached [Aster *et al.*, 2005]. One of the solutions to that problem is to repeat the optimization with various starting points and monitor if they all converge to the same solution [Mary *et al.*, 1998]. However, with the increase in model dimensions the gradient methods is prone to finding a local minimum closest to the starting point rather than the desired global minimum and it is, therefore, the method of choice only for optimizations of low number of parameters [Aster *et al.*, 2005].

To identify the global minimum, an exhaustive global search method that visits the entire model space using a random walk techniques based on Monte Carlo search have been developed.

This method take advantage of the fact that all local likelihood maxima are sampled if a sufficient number of iterations are performed; for such reason, this method avoid entrapment in local minima and is therefore ideally suited for highly non-linear optimization problems [Mosegaard and Tarantola, 1995]. Each optimization step is only dependent on the previous. The most efficient Monte Carlo method to sample the model space is the Metropolis algorithm [Metropolis *et al.*, 1953]. The Metropolis algorithm performs a random walk (comparable to a Brownian motion) in the model space. The walk is guided by probabilistic rules that decide if a move is accepted or rejected: moves that improve the fit are always accepted, while moves that degrade the fit are accepted with a certain probability. This allows the algorithm to move in and out of local minima.

An advantage of the Monte Carlo search with respect to gradient methods is that it is independent of the structure and analytical properties of the cost function and only requires the evaluation of the cost function. Also, the method is independent of the initial guess because the algorithm enforces a randomization of the initial guess. The major disadvantage is that the

stochastic nature of the algorithm requires a large amount of computer time to reach an acceptable final state. The factor controlling the numerical demands of MCMC are the complexity of the model which was integrated many thousand of times, and the number of parameters to optimize.

In the next section, we will describe the two different methods from a mathematical point of view. Following *Mosegaard and Tarantola* [1995], henceforth it is convenient to adopt the subsequent notation:

- s denotes the single station where observations are collected.
- t is the time (it assumes discrete values since there is a finite sampling time step).
- \mathbf{p} is the vector of parameters necessary within the model to make predictions.
- $\mathbf{M}(\mathbf{p}, s, t)$ is the prediction of the model, given a certain \mathbf{p} , at station s at time t .
- $\mathbf{D}(s, t)$ is the single observation taken at station s at time t .

In the following \mathbf{D} and $\mathbf{M}(\mathbf{p})$ and \mathbf{p} will be used as vectors.

3.2.1.1 CONJUGATE GRADIENT METHODS

The aim of an inverse algorithm is to find a global minimum of a cost function $J(\mathbf{p})$ in the multidimensional \mathbf{p} -space; in other words, we try to reduce the time-space misfits between the observed data \mathbf{D} and modeled variables $\mathbf{M}(\mathbf{p})$, with the constraints of the model equations and their parameters \mathbf{p} . In the Bayesian context, the cost function is usually the sum of two terms: the first term weights the uncertainties of the initial parameters with their respective *a priori* error covariances. The second term is the sum over time of all data-model misfits at observation locations, weighted by measurement error covariances. Generally, we can write the cost function as:

$$J(\mathbf{p}) = \frac{1}{2}[\mathbf{p} - \mathbf{p}_0] \mathbf{C}_0^{-1} [\mathbf{p} - \mathbf{p}_0]^T + \frac{1}{2}[\mathbf{M}(\mathbf{p}) - \mathbf{D}] \mathbf{C}_D^{-1} [\mathbf{M}(\mathbf{p}) - \mathbf{D}]^T \quad (3.1)$$

Here the upper index T indicates matrix transposition, \mathbf{p}_0 is the vector of a priori guesses for the parameters to optimize, \mathbf{C}_0 is the diagonal matrix of squared errors that we associate to the a priori guesses, and \mathbf{C}_D is the diagonal matrix of the squared errors that we associate to the observations.

The solution is sought iteratively by performing several evaluations of the cost function and of its gradient according to:

$$\nabla_{\mathbf{p}} J = 0 \quad (3.2)$$

Descent methods iteratively determine the descending directions along the cost function surface. The minimization can be stopped by limiting the number of iterations, or by requiring that the norm of the gradient $\|\nabla J(\mathbf{p})\|$ decreases by a predefined amount during the minimization: this value is an intrinsic measure of how much the analysis is closer to the optimum.

We employ two different methods to minimize the cost function: 1st derivative (or *gradient*) of $J(\mathbf{p})$ to model parameters \mathbf{p} $\left(\frac{\partial J(\mathbf{p})}{\partial \mathbf{p}} \right)$ yields direction of steepest descent. The method of steepest descent approaches the minimum in a zigzag manner, where the new search direction is orthogonal to the previous (*Figure 3.*).

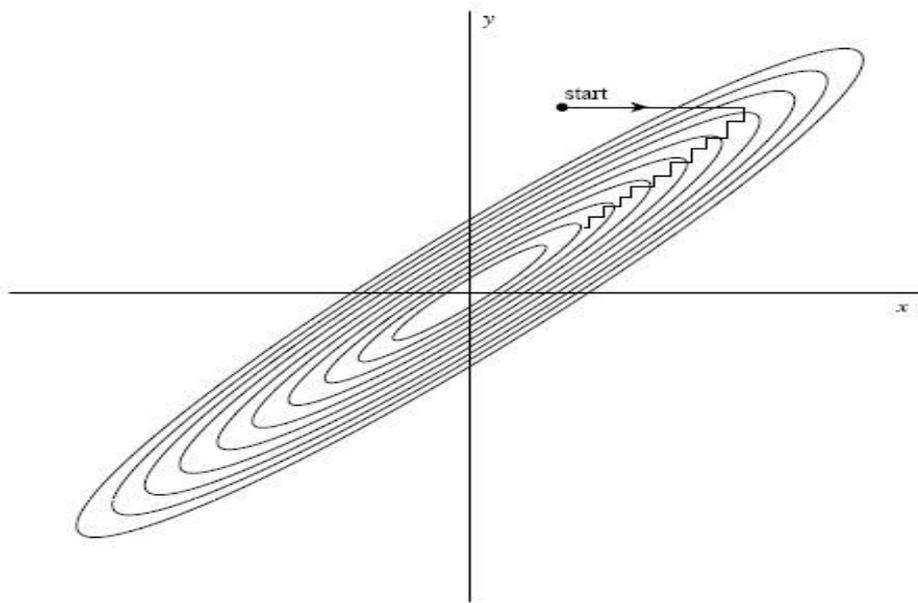


Figure 3.1. Successive minimizations along coordinate directions in a long, narrow “valley” (shown as contour lines). The step size gets smaller and smaller, crossing and recrossing the valley (from Press et al. [1992]).

The search starts at an arbitrary point X_0 and then slide down the gradient, until we are close enough to the solution. In other words, iteration continues until the extreme has been determined with a chosen accuracy ε .

Unless the valley is optimally oriented, this method is extremely inefficient, taking many tiny steps to get to the minimum, crossing and re-crossing the principal axis (*Figure 3.*) [*Press et al., 1992*].

2nd derivative (or *Hessian*) of $J(\mathbf{p})$ $\left(\frac{\partial^2 J(\mathbf{p})}{\partial(\mathbf{p})^2} \right)$ yields curvature of J . The product of the Hessian

by the local gradient yields a vector that directly points toward the minimum, even if this vector is oblique with respect to the direction of the steepest descent.

It is easy to see that the gradient (or Newton) methods correspond to obtaining at the current point \mathbf{p}_i the “paraboloid” that is tangent to the function $J(\mathbf{p})$ and that has the same local curvature and jumping to the point where this tangent paraboloid reaches its minimum [*Tarantola, 2005*].

Gradient-based methods try to find a compromise between two somewhat contradictory pieces of information. By definition, a gradient-based method uses *local* information on the function to be optimized, i.e., information that makes full sense in a small vicinity of the current point but does not necessarily reflect the properties of the function in a large domain. But each iteration of a gradient-based method tries to make a jump as large as possible, in order to accelerate convergence. For the intended finite jumps, the local information brought by the gradient may be far from optimal. In most practical applications, the user of a gradient-based method may use physical insight to ‘correct’ the gradient, in order to define a direction that is much better for a finite jump [*Tarantola, 2005*].

Choosing the right method to be used in an inverse problem is totally problem dependent, and it is very difficult to give any suggestion at the general level. For small-sized problems, Newton methods are easy to implement and rapid to converge. For really large-sized problems, the linear system that has to be solved in the Newton method may be prohibitively expensive and sometimes the methods may not work well [*Tarantola, 2005*].

Just to test the performances of the gradient methods a toy model has been developed; a toy model is a simplified set of objects and equations relating them that can nevertheless be used to understand a mechanism that is also useful in the full non-simplified theory.

The complexity of the equations built in the toy model handles both linear and non-linear systems, while the number of parameter to optimize have been changed from few parameters to a huge number according to the LPJ model complexity. In the toy model each number of parameters has associated a given equation (i.e. model, where $y=f(x)$). The data have been created as follow: once the number of parameters to optimize has been chosen we initialized the *true* parameter values to create the model solution, while the observation have been computed adding a Gaussian noise to the model solution values. The prior (initial) model estimates have been created using *guess* parameter values (different from *true* values). Once both observations

and prior model estimates have been collected the minimization starts; at the end of the process *posterior* parameter values have been compared with the true solution in order to check the efficiency and performances of the minimization algorithms.

Figure 3.2 shows results of the toy model in the case of 2, 3, 4, and 10 parameters respectively.

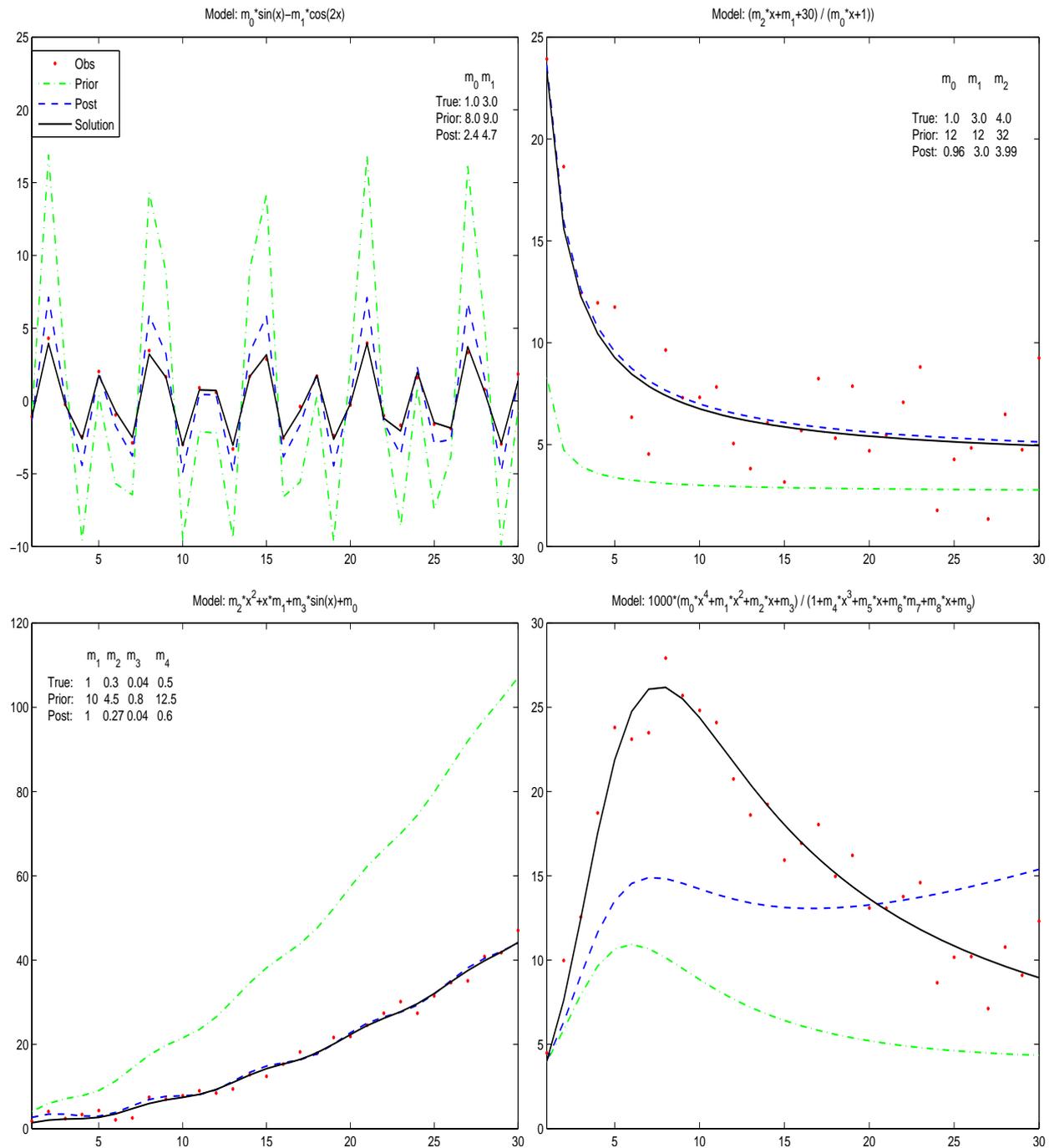


Figure 3.2. Different models used to check the performances of the gradient method. The panels show the toy model with 2, 3, 4, and 10 parameters respectively. Posterior parameters are compared to the known solution to check the reliability of the solution. The posterior values, however, are highly dependent on the prior guesses.

The posterior parameters are very close to the right solution when the models 3 and 4 have been taken into account, while results of the model with two parameters improves after optimization but posterior parameters are not so close to the correct solution. Finally, in the model with 10 parameters the optimization fails. The simulations we performed to check the validity of this method and this results showed us how this approach is too sensitive to the initial conditions: in fact, some repeated test allow us to get right values of all parameter in the models with 2, 3, 4, 5, and 6 but repeating optimization many times starting from different initial prior parameters.

Under these circumstances several hundreds of optimisations need to be performed in LPJ, starting from different initial parameter values in order to give reliable optimal estimates.

As a consequence of the results described above, we developed also a different optimization method. Moreover, in this work we plan to optimize a wide number of LPJ parameters, so, as already stated above, the gradient methods is not appropriate for such data assimilation. In the next sections will be described the *Metropolis algorithm* we applied in our inversion problem and we will refer only to this method in the whole work.

3.2.1.2 METROPOLIS ALGORITHM

In the MCMC context, the approach to inverse problems is characterized by a probabilistic definition of the solution [Tarantola, 2005]. In such case, the cost function $J(\mathbf{p})$ may be written in a similar way of the gradient method, i.e. as the sum of two terms, the first $J_0(\mathbf{p})$ weighing the a priori information, the second $J_D(\mathbf{p})$ weighing the misfit between model predictions and observations; these variables are related as described below:

$$J(\mathbf{p}) = J_0(\mathbf{p}) + J_D(\mathbf{p}) \quad (3.3)$$

$$J_0(\mathbf{p}) = \frac{1}{2}(\mathbf{p} - \mathbf{p}_0)^T \mathbf{C}_0^{-1}(\mathbf{p} - \mathbf{p}_0) \quad (3.4)$$

$$J_D(\mathbf{p}) = \frac{1}{2}[\mathbf{M}(\mathbf{p}) - \mathbf{D}]^T \mathbf{C}_D^{-1}[\mathbf{M}(\mathbf{p}) - \mathbf{D}] \quad (3.5)$$

On adopting the probabilistic viewpoint, no actual optimal \mathbf{p} exists as a constant, rather, \mathbf{p} must be regarded as a vector of aleatory variables having a joint probability distribution $\sigma(\mathbf{p})$. The goal

of the MCMC optimization is to sample enough often such probability distribution in order to get the statistical properties of posterior PDF (mean, median, mode, variance, skewness, and kurtosis).

In the Bayesian context the true joint probability distribution $\sigma(\mathbf{p})$ is generally written as [Mosegaard, 2006]:

$$\sigma(\mathbf{p}) = k\rho(\mathbf{p})L(\mathbf{p}) \quad (3.6)$$

where k is a normalization constant, while $\rho(\mathbf{p})$ and $L(\mathbf{p})$ are two multi-dimensional Gaussian probability distributions, defined as:

$$\rho(\mathbf{p}) = \exp[-J_0(\mathbf{p})] \quad (3.7)$$

$$L(\mathbf{p}) = \exp[-J_D(\mathbf{p})] \quad (3.8)$$

anyhow, this method works also for non-Gaussian distributions [Knorr and Kattge, 2005]. With these definitions we also simply have:

$$\sigma(\mathbf{p}) = k \exp[-J(\mathbf{p})] \quad (3.9)$$

The task is to retrieve sufficient information on the latter PDF. According to Mosegaard [1998] this inverse problem can most efficiently be solved by direct sampling of the PDF in the parameter space using the Metropolis algorithm consisting in an iterative algorithm that, starting from an initial guess for the parameter vector, generates a sequence of n successive points in the parameter space whose distribution, for n large enough, tend to reproduce the underlying true PDF of the aleatory vector variable \mathbf{p} .

Actually, not one but a certain number of point successions can be generated (in the present work we use six different sequences), each starting from a different initial point; the various initial points are spread around the prior guess-point at a mean distance given by the respective vector variance [Gelman *et al.*, 1995]. This avoid, to the six sequences, to start the exploration in the parameters space from the same initial point.

Within a given succession, the procedure to generate the next point is to move from the last point with a random step $\Delta\mathbf{p}$, chosen with a uniformly distributed random number generator [Press *et al.*, 1992] having mean zero and standard deviation set, separately for each parameter, to a

fraction (called step length factor) of the length of its prior standard deviation. If \mathbf{p}_j is the last point, the new point $\mathbf{p}_j + \Delta\mathbf{p}$ so generated, is accepted or rejected according to a 2-stage version of the Metropolis algorithm. In the first stage the step is rejected only if the random number generator, uniformly distributed between 0 and 1, draws a number $r > \rho(\mathbf{p}_j + \Delta\mathbf{p})/\rho(\mathbf{p}_j)$. In the second stage the step is rejected according to the same procedure as in the first stage, except that $\rho(\mathbf{p})$ is replaced by $L(\mathbf{p})$. Finally, the step is accepted if it overcomes both tests. In this case the next point in the succession will be $\mathbf{p}_{j+1} = \mathbf{p}_j + \Delta\mathbf{p}$, else the previous point is chosen again: $\mathbf{p}_{j+1} = \mathbf{p}_j$ [Knorr and Kattge, 2005; Braswell et al. 2005; Muller et al., 2007]. Note that only the second stage requires model execution; for such reason the first test is executed first. Figure 3.3 shows the flow diagram of MCMC.

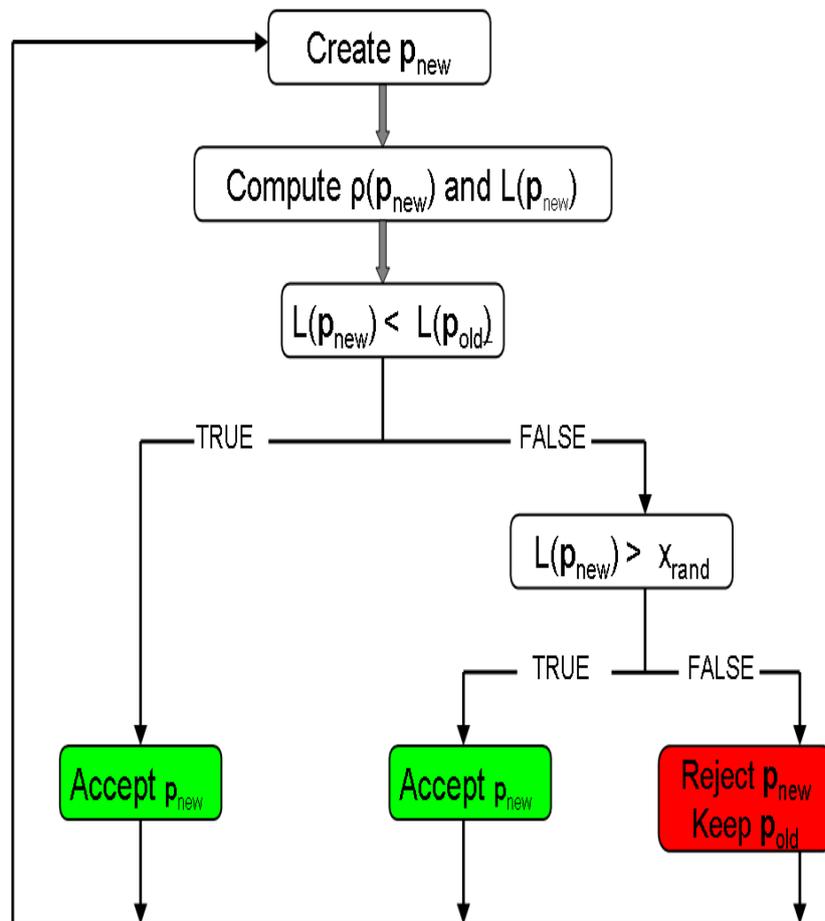


Figure 3.3. Flow diagram of the Markov chain Monte Carlo method using the Metropolis algorithm (see the above equations and the text for the explanation of the variables).

The sampling ends when the Gelman et al. [1995] convergence criterion is reached. With this criterion, the accuracy of the sampling can be monitored by a reduction factor R that declines to

1 for $n \rightarrow \infty$ [Knorr and Kattge, 2005; Braswell et al. 2005; Muller et al., 2007]. This criterion can be briefly stated as follows. Assigned one parameter, along a given sequence, after a conspicuous number of steps, at each successive step one can compute (better after discarding a certain number of iterations from the beginning of the random search) the intra-sequence variance, after having estimated the mean along the sequence as the arithmetic average over the residual sequence. Then, the mean of these intra-sequence variances, $\overline{s^2}$, may be computed. This quantity gives a measure of the mean tendency to oscillate around their own means of the various points belonging to the same sequence, while all the sequences tend to settle in the same spot in parameter space. On the other hand, one can simultaneously evaluate how much the along-sequence means deviate from their own mean. This within-sequence variability can be estimated by computing the variance of the along-sequence means, already calculated in the previous step. This quantity s_m^2 is an estimate of the variance of each along-sequence mean. If n is the length of each sequence, the latter variance must be close to the intra-sequence variance divided by n . Thus for n large enough, $\overline{s^2}$ and $n s_m^2$ tend to approximate the same quantity, with the second quantity exceeding the first, at least until all sequences have gathered onto the same spot in parameter space. As a consequence the statistics

$$\mathbf{R}_n = \sqrt{\frac{\frac{n-1}{n} \overline{s^2} + \frac{1}{n} (n s_m^2)}{\overline{s^2}}} = \sqrt{1 + \left(\frac{s_m^2}{\overline{s^2}} - \frac{1}{n} \right)} \quad (3.10)$$

assumes for every n a value greater than 1 and declines to 1 for $n \rightarrow \infty$. Note that there are as many \mathbf{R}_n as parameters. A value of $R < 1.2$ is acceptable for most cases and can serve as stop criteria for the optimization run [Knorr and Kattge, 2005; Braswell et al. 2005; Muller et al., 2007]. In the present work, we stopped the sampling once the reduction factors for all parameters had reached the value of 1.08.

The values at the beginning of the chain (10-50% is often found to be adequate) need to be discarded [Van Oijen et al., 2005]; they represent the “burn-in time”, i.e. the time required by the chain to adjust to the appropriate conditional distribution.

3.2.2 The optimization schemes

The MCMC procedure does not provide an analytical formula for the posterior probability distribution of the parameters, but a large sample of values from the posterior. So, once the search is stopped and a sufficiently long sequence of parameter values has been collected, the optimized parameter values may be taken from the mean, median or mode of each distribution, or we may simply choose the maximum a posteriori estimate (i.e., the value corresponding to the maximum of posterior distribution $\sigma(\mathbf{p})$) [Van Oijen et al., 2005; Trudinger et al., 2007]. Thus one has various estimation methods to choose the optimized parameter values.

On the other hand, there are various choices about the way of assimilating the data during an optimization. These alternatives depend on the data assimilation strategy. The observations \mathbf{D} , used to define $L(\mathbf{p})$, can be either a part or the whole of the available data. The observations data, used during our optimization process, consist in eddy covariance fluxes measured from 2000 to 2003. A detailed description of eddy fluxes will be given in the next sections. In the present work we followed two paths (i.e. strategies) to assimilate observations: first, we ran a separate optimization for each year (2000, 2001 and 2002) and we got, at the end of each simulation, a set of optimized parameters valid for that year. Next, we ran also a single optimization by assimilating the data of all the 3 years simultaneously (2000-2002). The data covering year 2003 have been used to check the results of optimization.

As for the estimation method, we chose to compute the mean and the maximum a posteriori estimate for either assimilation strategies. This make two different set of optimized parameters for each assimilation strategy; henceforth, these four schemes are so labeled:

- *Post Best*: Data assimilation is done year by year; the optimal values are those maximizing $\sigma(\mathbf{p})$. Then, we can average over the three years (2000, 2001 and 2002) to yields the final optimal parameter values (this will be discussed in the result section in detail).
- *Post Mean*: Data assimilation is year by year; the optimal values are estimated with the means computed using only the 2nd-half of n points for each sequence (Figure 3.4). Then averaging over the years yields the final optimal parameter values.
- *Post 3 Best*: Data assimilation uses all the 3-years data; the optimal values are those maximizing $\sigma(\mathbf{p})$.

- *Post 3 Mean*: Data assimilation uses all the 3-years data; the optimal values are estimated with the means computed using only the 2nd-half of n points for each sequence (*Figure 3.4*).

The mean and variance are unlikely to be sufficient for the posterior distribution of the parameters, unless the quantity of data is large. In this work, we therefore chose to evaluate model performance also with respect to the maximum a posteriori estimate of $\sigma(\mathbf{p})$ even if MCMC is a probabilistic method and it is not designed to provide an analytical value of the maximum of the cost function unlike gradient methods. In such case, the maximum a posteriori estimate may be thought of as a measure of the single “best” parameter value. *Figure 3.4* shows, how, for a given parameter (g_m , see *Table 3.1*), Metropolis algorithm explores randomly the parameter space starting from the six different initial-guess points dispersed around the prior value. It can be seen that, in order to reach the convergence a lot of iteration steps were necessary. The number of iterations needed to reach the convergence for MCMC algorithms depends on the complexity of the model (number of parameters to optimize, number and complexity of formulas that use these parameters, etc.) we are interested to optimize.

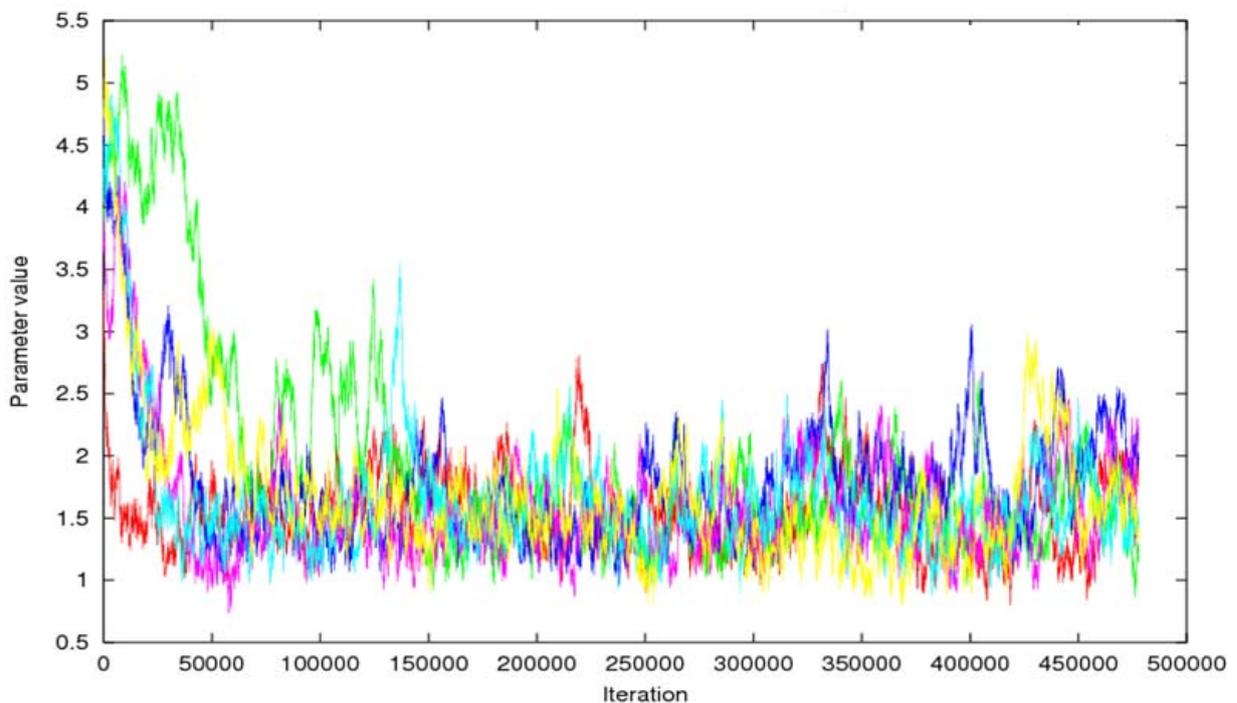


Figure 3.4. Given a model parameter (g_m , see *Table 3.1*), Metropolis algorithm explores randomly the parameter space starting from six different points; each color represents one of the six different sequences. Generally, only the second half of the succession of points (namely after an epoch when it is supposed that the search/sampling procedure has finally become stationary) is used to estimate the statistical moments of the PDF (mean, mode, median, covariances and percentiles).

A major advantage of the MCMC technique is that it samples the posterior PDF without any assumption about its shape. The shape of marginal distributions of single parameters can be nicely visualized by histograms. In the present work MCMC yields a sample with a normal or quasi-normal distribution for all parameters taken into account during the optimization; in *Figure 3.5* we show, as example, the posterior marginal distribution of the parameter bc_3 (see *Table 3.1*).

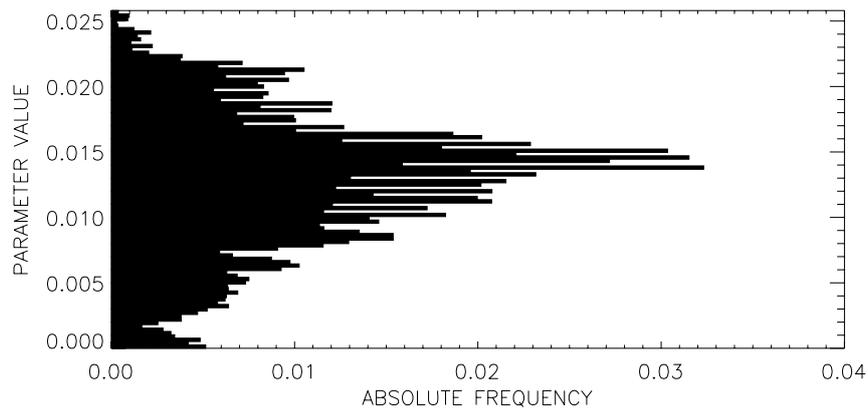


Figure 3.5. Example of the posterior marginal distribution for the parameter bc_3 (see *Table 3.1*)

3.2.3 Model description and input data

The LPJ-DGVM [Sitch *et al.*, 2003] is a process-based model representing key ecosystem processes governing terrestrial biogeochemistry and biogeography. LPJ simulates the water and carbon exchange between biosphere and atmosphere by means of a given set of parameters and input variables. For each grid cell, vegetation is described in terms of the Fractional Percentage Coverage (FPC) of 9 different Plant Functional Types (PFTs) that are able to compete for space and resources [Sitch *et al.*, 2003]. Seven PFTs are woody (three temperate, two tropical and two boreal) and two herbaceous (*Table 3.1*); the PFTs have different photosynthetic (C_3 , C_4), phenological (deciduous, evergreen), and physiognomic (tree, grass) attributes [Sitch *et al.*, 2003]. The PFTs may in principle co-exist at any location, depending on plant competition and a set of environmental constraints. Their relative proportion is determined by competition among types with typical ecological strategies for dealing with temperature, water and light stress. Dispersal processes are not explicitly modelled and an individual PFT can invade new regions if its bioclimatic limits and competition with other PFTs allow establishment. *Figure 3.6* shows the model logic.

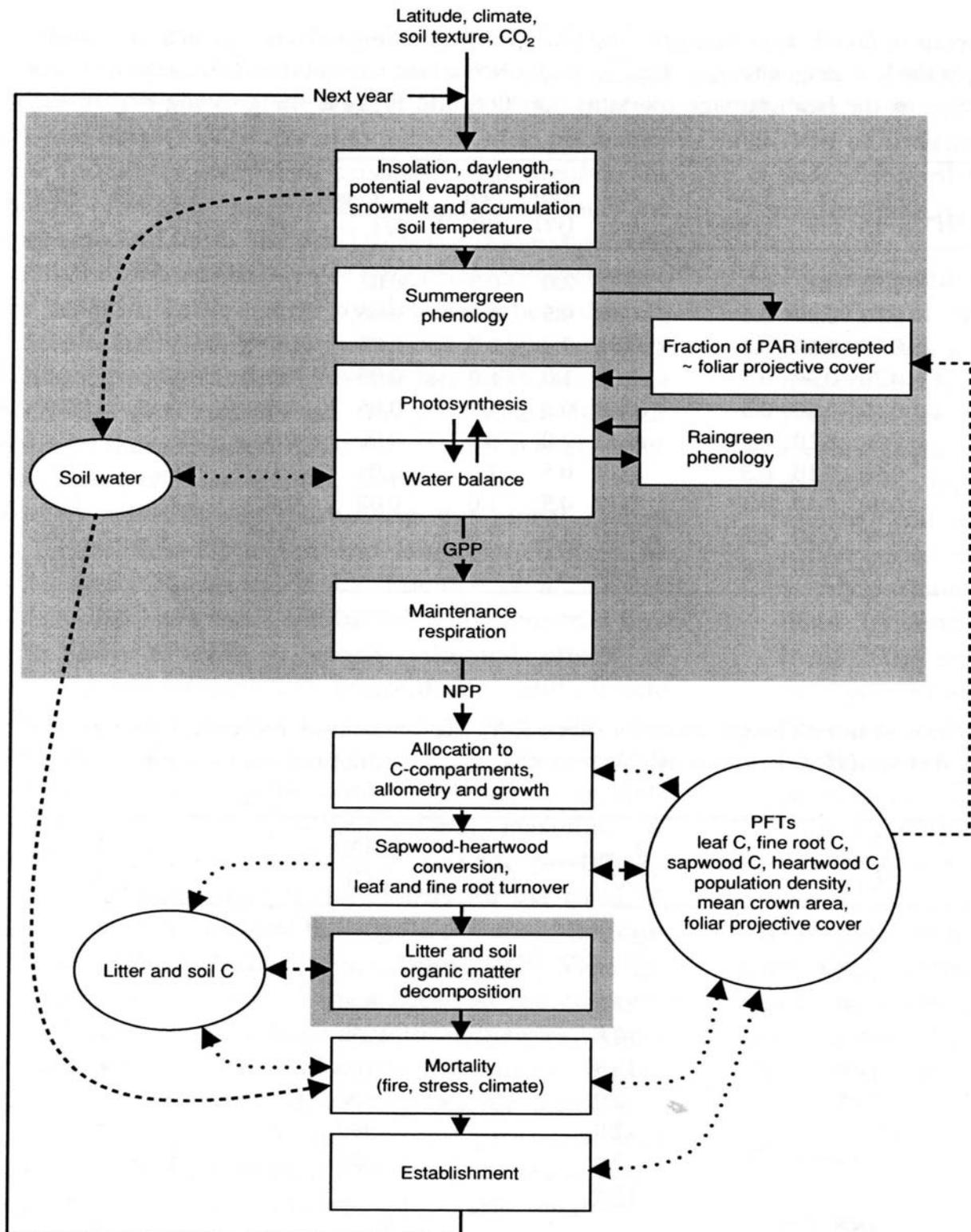


Figure 3.6. A flowchart describing the order individual process (in boxes) performed in all grid cells during one simulation year. The dashed lines represent exchange of information between vegetation and soil state variables and the individual modules, with arrow representing the direction of information flow. Modules with a shaded background are called on a daily or monthly time step, the remainder are called annually (from Sitch et al. [2003]).

PFT	z_1 (-)	z_2 (-)	g_{min} (mm/s)	r_{fire} (-)	a_{leaf} (yr)	f_{leaf} (yr ⁻¹)	$f_{sapwood}$ (yr ⁻¹)	f_{root} (yr ⁻¹)	$t_{mort,min}$ (°C)	S_{GDD} (°C)
Tropical Broadleaved Evergreen	0.85	0.15	0.5	0.12	2.0	0.5	0.05	0.5	-	-
Tropical Broadleaved Raingreen	0.7	0.3	0.5	0.5	0.5	1.0	0.05	1.0	-	-
Temperate Needleleaved Evergreen	0.6	0.3	0.3	0.12	2.0	0.5	0.05	0.5	-	-
Temperate Broadleaved Evergreen	0.7	0.3	0.5	0.5	1.0	1.0	0.05	1.0	-	-
Temperate Broadleaved Summergreen	0.7	0.2	0.5	0.12	0.5	1.0	0.05	1.0	-	200
Boreal Needleleaved Evergreen	0.9	0.1	0.3	0.12	2.0	0.5	0.05	0.5	23	-
Boreal Broadleaved Summergreen	0.9	0.1	0.3	0.12	0.5	1.0	0.05	1.0	23	200
C ₃ Perennial Grass	0.9	0.1	0.5	1.0	1.0	1.0	-	0.5	-	100
C ₄ Perennial Grass	0.9	0.1	0.5	1.0	1.0	1.0	-	0.5	-	100

Table 3.1. Specific PFT parameter values: z_1 and z_2 are the fraction of fine roots in the upper and lower soil layers, respectively; g_{min} is the minimum canopy conductance; r_{fire} is the fire resistance; a_{leaf} is the leaf longevity; f_{leaf} , $f_{sapwood}$ and f_{root} are the leaf, sapwood and fine root turnover times, respectively; $t_{mort,min}$ is the temperature base in the heat damage mortality function and S_{GDD} is the growing degree day requirement to grow full leaf coverage.

In addition to the parameters controlling physiology and dynamics, each PFT has associated specific parameters which determine whether it can survive and or regenerate under the climate conditions prevailing in a given grid cell at particular time during the simulation (Table 3.2) [Sitch *et al.*, 2003].

PFT		W/H*	$T_{c,min}$ (°C)	$T_{c,max}$ (°C)	GDD_{min} (°C)
Tropical Broadleaved Evergreen	(TrBE)	W	15.5	-	-
Tropical Broadleaved Raingreen	(TrBR)	W	15.5	-	-
Temperate Needleleaved Evergreen	(TNE)	W	-2.0	22.0	900
Temperate Broadleaved Evergreen	(TBE)	W	3.0	18.8	1200
Temperate Broadleaved Summergreen	(TBS)	W	-17.0	15.5	1200
Boreal Needleleaved Evergreen	(BNE)	W	-32.5	-2.0	600
Boreal Broadleaved Summergreen	(BBS)	W	-	-2.0	350
C ₃ Perennial Grass	(C3)	H	-	15.5	-
C ₄ Perennial Grass	(C4)	H	15.5	-	-

* W = Woody; H = Herbaceous.

Table 3.2. Specific PFT bioclimatic limits: $T_{c,min}$ = minimum coldest-month temperature for survival; $T_{c,max}$ = maximum coldest-month temperature for establishment; GDD_{min} = minimum degree-day sum (5 °C base) for establishment.

The simulations are driven by monthly fields of mean temperature, precipitation and cloud cover taken from the Climate Research Unit (CRU) monthly climate dataset on a 0.5° x 0.5° global grid [Mitchell and Jones, 2005]. An annual data set of historical global atmospheric CO₂

concentrations, extending from 1901 to 2003, was obtained from *Keeling and Whorf* [2005]; these data were derived from a combination of ice-core measurement and atmospheric observations [*Sitch et al.*, 2003]. Soil texture data were based on the FAO soil dataset [*Zobler*, 1986].

Monthly fields are linearly interpolated to obtain daily values, directly used at daily time steps to simulate short term processes, such as photosynthesis, respiration and evaporation. A disadvantage of using monthly climate data interpolated to daily values is that the response of ecosystem processes to daily variation is not considered. If these responses are non-linear, cumulative effects of ignoring daily variation may also affect the results on a monthly and annual time scale.

Photosynthesis is a function of Absorbed Photosynthetically Active Radiation (APAR), temperature, atmospheric CO₂ concentration, day length, canopy conductance, and biochemical pathway, using a form of the Farquhar scheme [*Farquhar et al.*, 1980; *Collatz et al.*, 1992], with leaf-level optimized nitrogen allocation [*Haxeltine and Prentice*, 1996] and an empirical convective boundary layer parameterization [*Monteith*, 1995] to couple carbon and water cycles [*Gerber et al.*, 2004].

Evapotranspiration is calculated as the minimum of a plant and soil limited supply function, proportional to the soil moisture in the rooting zone, and the atmospheric demand [*Haxeltine and Prentice*, 1996, *Sitch et al.*, 2003]. Both evapotranspiration and photosynthesis are calculated daily.

Maintenance respiration (R_m) is calculated as function of C/N, temperature, tissue biomass and phenology [*Farquhar et al.*, 1980; *Haxeltine and Prentice*, 1996]. After maintenance respiration is subtracted from GPP, 25% is taken as growth respiration (the cost of producing new tissues), while annual net primary production is:

$$NPP = 0.75(GPP - R_m) \quad (3.11)$$

The assimilated carbon is allocated annually to four pools (leaves, sapwood, heartwood and fine roots) to satisfy a set of allometric and functional relations [*Sitch et al.*, 2003]. Leaf and root turnover, as well as plant mortality, redistribute carbon to a litter and to a slow and a fast soil carbon pool. Each PFT has associated above-ground and below-ground litter pool (*Figure 3.7*). Decomposition rates of soil and litter organic carbon depend on soil temperature [*Lloyd*, 1994] and moisture [*Foley*, 1995]. The 70% of the decomposed litter goes directly into atmosphere as CO₂ while the remainder (30%) enters in the soil pool (Φ_{litter}). The 98.5% of reminder goes to

intermediate soil organic matter (Φ_{interm}) and 1.5% enters in the slow soil pool (Φ_{slow}). The sum of litter and soil decomposition terms gives total heterotrophic respiration R_h :

$$R_h = \Phi_{litter} + \Phi_{interm} + \Phi_{slow} \quad (3.12)$$

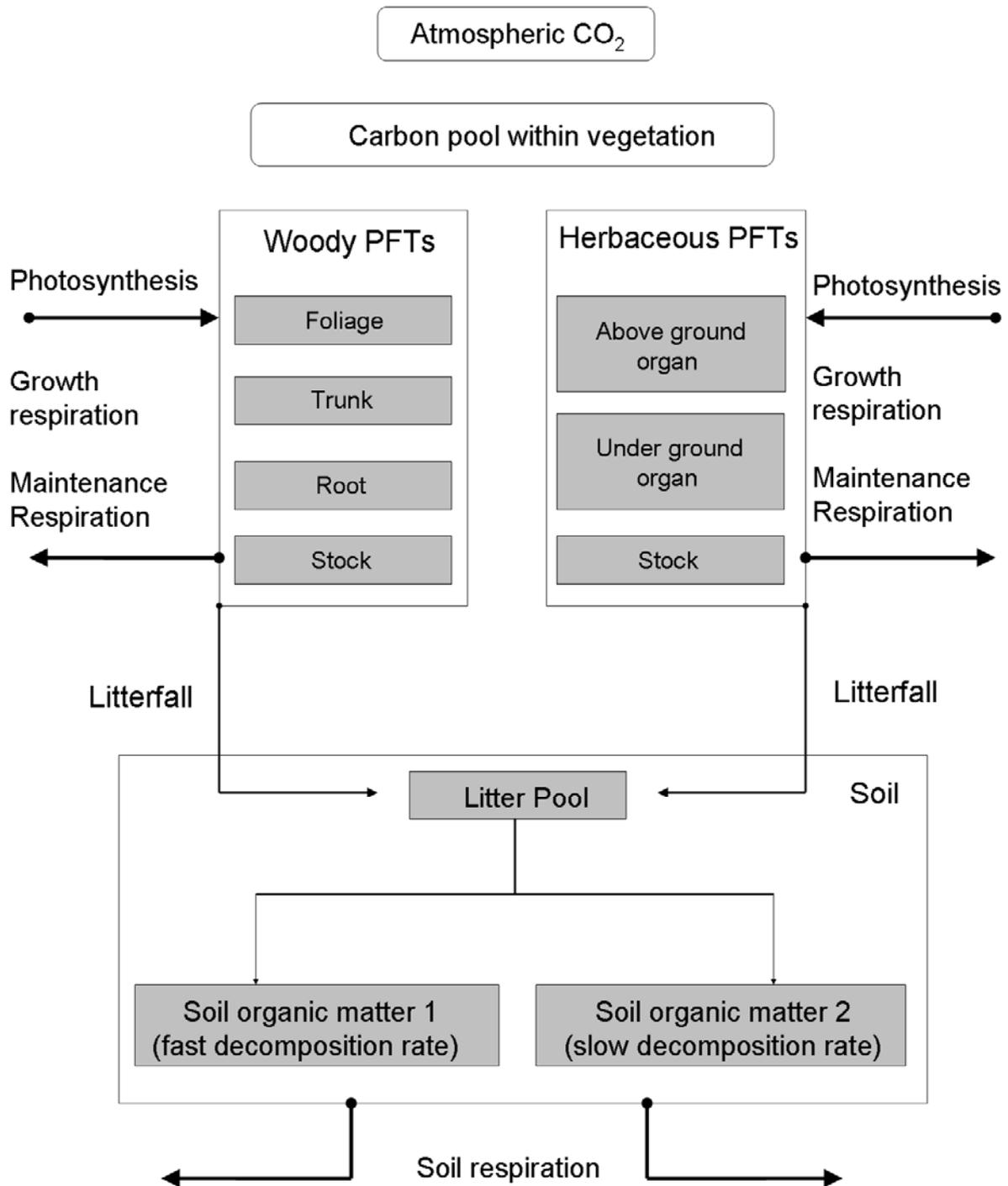


Figure 3.7. The carbon cycles as simulated in LPJ. Photosynthesis, autotrophic, and heterotrophic respiration are computed daily for each PFT.

Net ecosystem exchange, which represents the carbon balance at a point scale, is the difference between heterotrophic respiration and NPP, i.e.

$$NEE = R_h - NPP \quad (3.13)$$

Negative fluxes denote a net carbon flux from the atmosphere to the terrestrial biosphere (in such case we have a sink of carbon). The net carbon exchange of terrestrial ecosystems is the result of a delicate balance between uptake (photosynthesis) and losses (respiration), and shows a strong diurnal, seasonal, and inter-annual variability. Under favorable conditions, during daytime the net ecosystem flux is dominated by photosynthesis, while at night, and for deciduous ecosystems in leafless periods, the system loses carbon by respiration. The influence of climate and phenology can in some cases shift a terrestrial ecosystem from a sink to a source of carbon.

A typical simulation with LPJ starts from bare ground (no plant biomass present for each grid cell) and spins up for 1000 years until approximate equilibrium is reached with respect to carbon pools and vegetation cover. Usually, during the spin up phase, the model is driven with an approximately constant climate; however, since fires in many region occur only in drier years, repeating a single year's climate or even using a long-term climate average, can lead to anomalous results [Sitch *et al.*, 2003]. The spin up therefore requires the use of an interannually varying climate, with annual-average temperatures, precipitation and cloudiness fluctuation about constant long-term means [Sitch *et al.*, 2003].

In order to enhance computational efficiency, the size of the slow Soil Organic Matter (SOM) pool is solved analytically after 400 years. Then, the spin up is continued for a further 600 years to ensure closeness to equilibrium in the slower carbon pools. After the spin up, the model can be driven with a transient climate [Sitch *et al.*, 2003].

In the present work, LPJ have been initialized at site point with bare ground and spin up for 1000 years to achieve equilibrium in terms of preindustrial stable vegetation structure and carbon pools. During the spin up phase 30 years of varying climate from the beginning of the CRU [Mitchell and Jones, 2005] climate dataset were repeated continuously with preindustrial atmospheric CO₂ content. Starting from this equilibrium, the model was driven using the CRU climatology till 2000 using observed atmospheric CO₂ content. For each site we prescribed constant dominant PFT and FPC [Sitch *et al.*, 2003] by means of maximum woody LAI as measured at the site.

The implementation of the algorithms for the data assimilation inside the model required a lot of time and an extensive development of the code; in *Figure 3.8* is shown the original model tree functions and the new structure built in LPJ and used for the data assimilation.

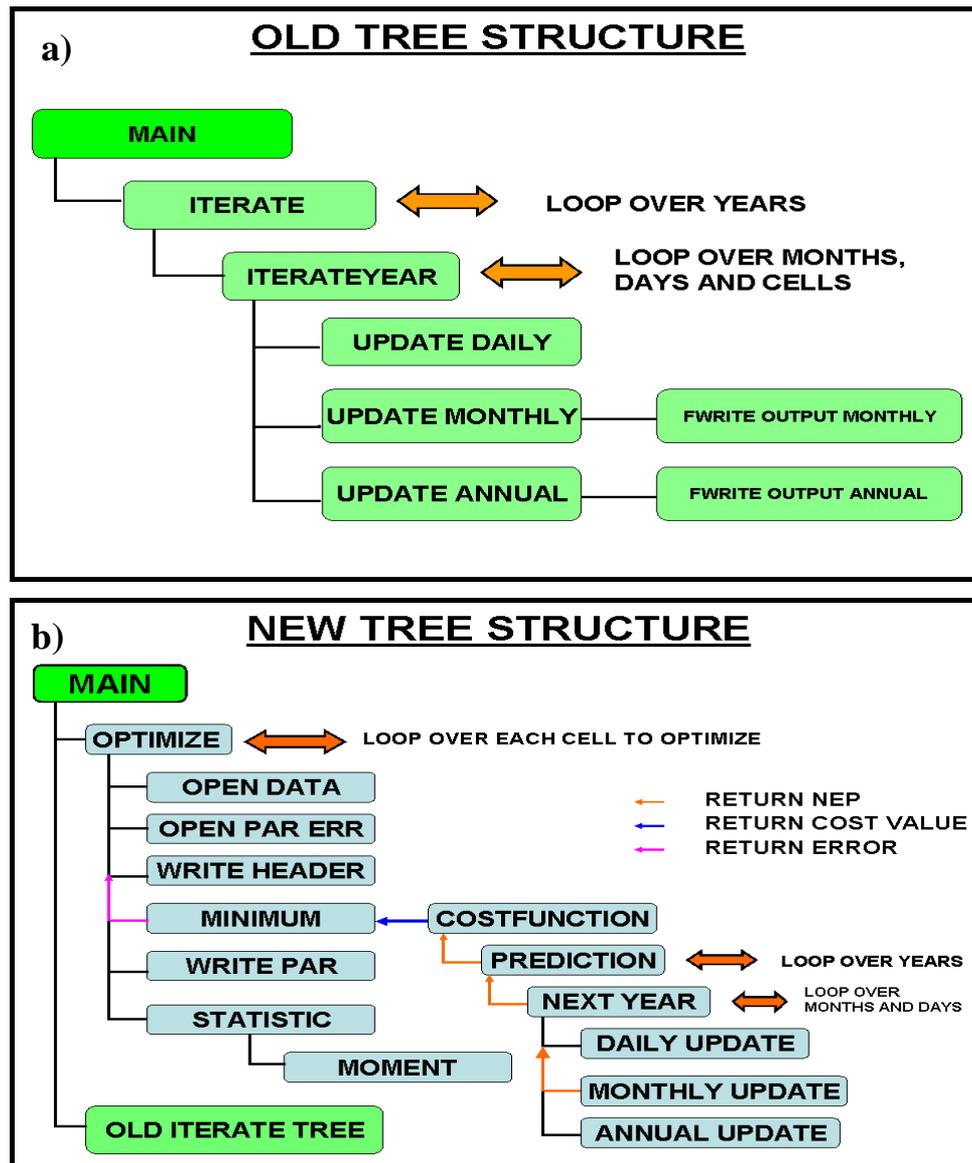


Figure 3.8. Original tree structure (a) and new functions (b) built inside the model.

3.2.4 Initial model parameters values and prior uncertainties

We chose the parameter to optimize from a deep ecological and biophysical literature search and from the knowledge of the most important parameters contributing to overall model uncertainty [Zaehle *et al.*, 2005]. In order to assess the parameter-based uncertainty on the global scale, Zaehle *et al.* [2005] attempted to identify the most important parameters contributing to the overall model uncertainty, and performed a global-scale uncertainty analysis with a set of 36 parameters (Table 3.3). The parameter importance for all 36 parameters was determined at a set of locations (81 sites) spanning all major biomes.

Parameter	Standard	Min	Max	Description	
<i>Photosynthesis</i>	θ	0.7	0.2	0.996	Co-limitation shape parameter
	k_{beer}	0.5	0.4	0.7	Extinction coefficient
	α_a	0.4	0.3	0.7	Scaling parameter (leaf to canopy)
	λ_m	0.8	0.6	0.8	Optimal c_i/c_a for C ₃ plants
	α_{C3}	0.08	0.02	0.125	Intrinsic quantum efficiency of CO ₂ uptake in C ₃ plants
	b_{C3}	0.015	0.01	0.021	Leaf respiration as a fraction of Rubisco capacity in C ₃ plants
	$\lambda_{m,C4}$	0.4	0.31	0.4	Optimal c_i/c_a for C ₄ plants
	α_{C4}	0.053	0.3	0.054	Intrinsic quantum efficiency of CO ₂ uptake in C ₄ plants
	b_{C4}	0.035	0.00205	0.0495	Leaf respiration as a fraction of Rubisco capacity in C ₃ plants
<i>Respiration</i>	r_{maint}	0.0495	0.066	0.0825	Tissue respiration rate at 10 °C
	r_{growth}	0.25	0.15	0.4	Growth respiration per unit NPP
	e_a	308.56	275	325	Activation energy for respiration
<i>Hydrology</i>	z_1	0.667	0.5	0.9	Fraction of fine roots in the upper soil layer
	E_{max}	5.0	2.4	6.2	Maximum daily transpiration rate
	g_m	5.0	2.5	18.5	Maximum canopy conductance
	g_{min}	0.3/0.5	0.22/0.42	0.38/0.58	Minimum canopy conductance
	α_m	1.4	1.1	1.5	Empirical evapotranspiration parameter
	$loss_{int}$	0.2/0.6	0.15/0.45	0.25/0.75	Interception loss parameter
<i>Allometry</i>	k_{allom1}	100	75	125	Used to compute the crown area
	k_{allom2}	40	30	50	Used to compute the height
	k_{allom3}	0.5	0.5	0.8	Used to compute the height
	$k_{la:sa}$	8000	2000	8000	Leaf to sapwood area ratio
	k_{rp}	1.6	1.33	1.6	Used to compute the crown area
	CA_{max}	15	7.5	30	Maximum woody PFT crown area
<i>Vegetation dynamics</i>	k_{mort1}	0.01	0.005	0.1	Asymptotic maximum mortality rate
	k_{mort2}	35	20	50	Growth efficiency mortality scalar
	est_{max}	0.24	0.05	0.48	Maximum sapling establishment rate
<i>Fire dynamics</i>	r_{fire}	0.12/0.5/1	-10%	+10%	Fire resistance
	m_e	0.3	0.0225	0.375	Litter moisture of extinction
	$fuel_{min}$	0.2	0.01	0.4	Minimum fuel load for fire spread
<i>PFT characteristics</i>	$f_{sapwood}$	0.05	0.01	0.2	Sapwood turnover rate
	α_{leaf}	0.5/1/2	0.11	8.0	Leaf longevity
	$dens_{wood}$	200	180	220	Specific wood density
<i>SOM pools</i>	τ_{litter}	2.85	1.23	5.26	Litter turnover time at 10 °C
	f_{air}	0.7	0.5	0.9	Fraction of the decomposed litter emitted as CO ₂ to the atmosphere
	f_{inter}	0.985	0.85	0.99	Fraction of soil-bound decomposed litter entering in the intermediate soil pool

Table 3.3. Key LPJ parameters used in Zaehle et al. [2005] to perform the uncertainty analysis: Standard value, literature range and parameter description. Table redrawn from Zaehle [2005].

The parameters having the greatest influence on the ecosystem carbon cycling as well as being of substantial importance in modeling vegetation dynamics are ranked in *Table 3.4*. The most important parameters controlling a given variable have been computed by means the Partial Correlation Coefficient (PCC): it has been calculated to identify the relative importance of the uncertainty contribution of a particular parameter to the total model output uncertainty [Zaehle, 2005].

The concept of partial correlation allows the identification of the contribution of a parameter x_j to the variation in the output variable y when the variation in y is influenced by a number of co-varying factors, or parameters, x_p . Since PCC relies on linear regression, it provides a measure for the linear relationship between x_j and y with the linear effects of all other parameters x_p removed. To identify potential non-linear monotonic effects Zaehle *et al.* [2005] used a rank-transformed input variables (RPCC, Conover and Iman [1981]).

Rank	GPP		NPP		R_h		Vegetation C		Litter C	
	Par.	RPCC	Par.	RPCC	Par.	RPCC	Par.	RPCC	Par.	RPCC
1	α_{c3}	0.864	α_{c3}	0.864	α_{c3}	0.801	α_{c3}	0.607	α_{c3}	0.791
2	α_a	0.733	α_a	0.704	α_a	0.662	$f_{sapwood}$	0.536	τ_{litter}	0.742
3	θ	0.513	θ	0.474	g_m	0.467	k_{mort1}	-0.459	α_a	0.593
4	g_m	0.463	g_m	0.463	θ	0.429	$k_{la:sa}$	-0.398	θ	0.405
5	b_{c3}	-0.309	r_{growth}	-0.297	k_{beer}	-0.303	g_m	0.379	g_m	0.397
6	k_{beer}	-0.249	k_{beer}	-0.27	r_{growth}	-0.255	α_a	0.37	$f_{sapwood}$	0.321
7	λ_m	0.149	b_{c3}	-0.268	b_{c3}	-0.242	est_{max}	-0.318	r_{growth}	-0.252
8	E_{max}	0.143	$f_{sapwood}$	0.23	α_{leaf}	0.201	θ	0.257	b_{c3}	-0.224
9	α_m	-0.127	a_{leaf}	-0.217	$f_{sapwood}$	-0.192	k_{mort2}	0.186	α_{leaf}	-0.193
10	α_{leaf}	-0.089	r_{maint}	-0.134	α_m	0.112	k_{allom2}	0.171	E_{max}	0.149
11	z_1	-0.077	λ_m	0.132	r_{maint}	-0.11	z_1	-0.149	k_{beer}	-0.146
12	$k_{la:sa}$	0.005	E_{max}	-0.128	λ_m	-0.109	b_{c3}	-0.143	r_{maint}	-0.121

Table 3.4. The twelve most important parameters controlling carbon fluxes and pools (taken from Zaehle [2005] and Zaehle *et al.* [2005]). The ranking was performed according to the average RPCC across all 81 grid cells. Regionally, the importance ranking may vary, as discussed in the text.

Although these parameters have a similar effect on A and NPP worldwide, differences in parameter importance can be observed in water-limited regions (e.g. Mediterranean sites) as characterised by a low ratio of actual to potential evapotranspiration. In such regions, parameters controlling plant water balance, i.e. g_m , E_{max} and α_m are relatively important [Zaehle *et al.*, 2005].

On the basis of the above results and the knowledge of the model formulation, we chose to constrain 12 parameters controlling directly or indirectly gross primary production and evapotranspiration: 5 of these are used to compute photosynthesis, 3 to compute water stress and water balance and the other 4 are specific PFT parameters. All the model parameters values used during data assimilation are listed in *Table 3.5* and *Table 3.6*.

Function	Name	Prior	Min	Max	Reference
<u>Photosynthesis</u>	b_{C3}	0.015	0.01	0.21	<i>Farquhar et al.</i> [1980]
	θ	0.7	0.2	0.996	<i>McMurtrie and Wang</i> [1993]
	α_a	0.5	0.3	0.7	<i>Haxeltine and Prentice</i> [1996]
	α_{C3}	0.08	0.02	0.125	<i>Collatz et al.</i> [1991]
	λ_m	0.8	0.6	0.8	<i>Haxeltine and Prentice</i> [1996]
<u>Water Balance</u>	α_m	1.391	1.1	1.5	<i>Monteith</i> [1995]
	g_m	3.26	2.5	18.5	<i>Magnani et al.</i> [1998]
	PT	1.32	1.1	1.4	<i>Monteith</i> [1995]

Table 3.5. Standard value, literature range and references of the LPJ model parameters used during inversion against eddy covariance data. Some parameter values may differ from *Table 3.3* because we used a more recent model version to perform the simulations.

Note that when we optimize g_{min} (*Table 3.6*) we are constraining two different parameters: the first one controls the minimum canopy conductance for trees while the last is related to C_3 grass. This depends on the prescribed vegetation: in fact, for each site, we prescribed the dominant trees, in terms of woody PFT, and the understory as C_3 grass.

PFT	z_1	E_{max}	g_{min}
Temperate needleleaved evergreen tree	0.6	5.0	0.3
Temperate broadleaved evergreen tree	0.7	5.0	0.5
Temperate broadleaved summergreen tree	0.7	5.0	0.5
Boreal needleleaved evergreen tree	0.9	5.0	0.3
C_3 grass	Not Optimized	Not Optimized	0.5

Table 3.6. Specific PFT parameters used during inversion against eddy covariance measurements; z_1 is the fraction of fine roots in the upper soil layer; E_{max} represents the maximum transpiration rate and g_{min} is the minimum canopy conductance.

The parameter θ [McMurtrie and Wang, 1993] is used in the daily net photosynthesis account to give a gradual transition between two limiting rates (*see Appendix*): the first one describes the response of photosynthesis to APAR, while the second one describes the Rubisco limited rate of photosynthesis [Haxeltine and Prentice, 1996]. The empirical parameter α_a [Haxeltine and Prentice, 1996] accounts for reduction in photosynthetically active radiation (PAR) utilization efficiencies in natural ecosystems, while a_{C3} [Collatz *et al.*, 1991] is the intrinsic quantum efficiency for CO₂ uptake as measured in laboratory studies. The quantum efficiency describes the amount of photosynthesis that occurs per unit number of photons adsorbed. In general, a_{C3} should be considered as valid and α_a constrained by flux tower data since the latter one represents the scaling. In fact, the scaling parameter α_a for quantum efficiency for CO₂ uptake is used to scale down laboratory-measured values of quantum efficiency to be more in accord with quantum efficiencies measured in the field. However, the stratification of species into plant functional types makes important to also investigate the optimal value of a_{C3} which is thought to being an average for different species.

The parameter b_{C3} [Farquhar *et al.*, 1980] is used to compute leaf respiration (R_d) as a function of Rubisco capacity (V_m) following Haxeltine and Prentice [1996]:

$$R_d = b_{C3}V_m \quad (3.14)$$

Instead of prescribing values for Rubisco capacity, the photosynthesis scheme uses the above equation in an optimization algorithm to predict the monthly values for Rubisco capacity that give the maximum non water-stressed daily rate of net photosynthesis (*see Appendix*).

Because of many observations, valid for C₃ species under non-water-stressed conditions, have shown that stomata respond in a way that maintains a constant ratio of intercellular (p_i) to ambient (p_a) CO₂ partial pressure of 0.6-0.8 [Haxeltine and Prentice, 1996], the model use $\lambda_m \leq 0.8$ to describe the relation $p_i = \lambda_m p_a$ under non-water-stress conditions. The previous equation show how λ_m is used to describe the biochemical dependence of daily net photosynthesis on internal CO₂ pressure and environmental variables.

As already introduced in the above paragraph, the evapotranspiration is calculated for each PFT as the minimum of a supply function (S), proportional to the soil moisture in the rooting zone (W_r) by means the maximum transpiration rate E_{max} (Table 3.6) [Haxeltine and Prentice, 1996; Sitch *et al.*, 2003]:

$$S = E_{max}W_r \quad (3.15)$$

and a demand function (D) calculated as a function of the potential canopy conductance (g_{pot}) following Monteith's [Monteith, 1995; Haxeltine and Prentice, 1996, Sitch et al., 2003] empirical relation between evaporation efficiency and surface conductance:

$$D = E_q \alpha_m \left[1 - \exp\left(\frac{-g_{pot}}{g_m}\right) \right] \quad (3.16)$$

where E_q is the equilibrium evapotranspiration rate calculated from latitude, temperature and sunshine hours as described by Haxeltine and Prentice [1996]. Finally, the Priestley-Taylor coefficient is proportional to α_m as discussed by Monteith [1995].

Since in LPJ model carbon and water fluxes are highly coupled through the canopy conductance, we chose to apply a multiple constraint approach that simultaneously optimize parameters controlling both assimilation and evapotranspiration.

The assignment of uncertainties to prior parameters play a pivotal role in the Bayesian inversion [Santaren et al., 2007]. Generally if hesitation exists in choosing the a priori uncertainty bars, it is of course best to be over-conservative and to choose them large [Tarantola, 2005]; in the absence of better knowledge we therefore assign a prior relative uncertainty equal to 0.15 in terms of 1 standard deviation to all parameters, except for α_a where it was assumed to be 0.18. Covariances among prior parameters are not considered.

3.2.5 Eddy-covariance flux data

The Eddy Covariance technique (EC) is a prime atmospheric flux measurement method to measure and calculate vertical turbulent fluxes within atmospheric boundary layers. It is a statistical method used in meteorology and other applications that analyzes high-frequency wind and scalar atmospheric data series, and yields values of fluxes of these properties.

The data consist of the vertical flux densities of CO₂, latent heat and sensible heat measured using sensors placed above the canopy, together with meteorological conditions (Figure 3.9).

The EC provides the only direct measurement of net terrestrial ecosystem-atmosphere CO₂ exchange, and is therefore an indispensable tool for understanding and monitoring the global carbon cycle [Friend et al., 2006].

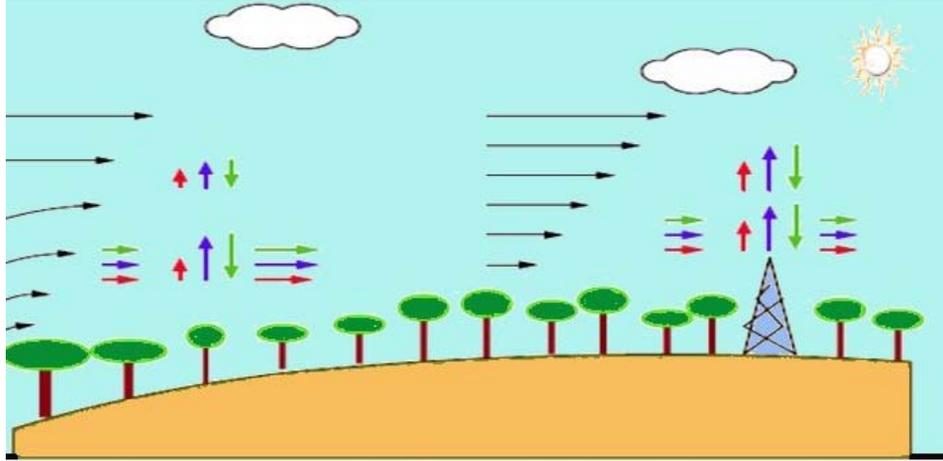


Figure 3.9. The vegetation exchanges with the atmosphere heat (red arrows), water (blue arrows) and CO_2 (green arrows); these fluxes are transported by the wind (black arrows) and we can measure such fluxes by means instruments placed above the canopies.

The fluxes are typically integrated to half-hourly or hourly means [Aubinet *et al.*, 2000]. Data gaps, caused by sensor failures or unsuitable micrometeorological conditions (e.g. heavy rainfall) are unavoidable [Falge *et al.*, 2001], but these can be filled using a variety of techniques to produce a continuous time series [Papale *et al.*, 2006].

Briefly, the EC is based on the conservation equation of a scalar (e.g. net carbon flux); if ρ_s is the scalar density, we can write the equation as [Valentini, 2003]:

$$\frac{\partial \rho_s}{\partial t} + u \frac{\partial \rho_s}{\partial x} + v \frac{\partial \rho_s}{\partial y} + w \frac{\partial \rho_s}{\partial z} = S + D \quad (3.17)$$

where u , v , and, w are the wind velocity components, respectively, in the directions of the mean (x), and lateral wind (y), and normal to the surface (z). S is the source/sink term and D is the molecular diffusion term. The lateral gradients and the molecular diffusion will be neglected afterwards. After application of the Reynolds decomposition ($u = \bar{u} + u'$; $v = \bar{v} + v'$; $w = \bar{w} + w'$; $\rho_s = \bar{\rho}_s + \rho_s'$, where the over bars characterize time averages and the primes fluctuations around the average), averaging, integration along z , and assumption of no horizontal eddy flux divergence, equation (3.18) becomes:

$$\int_0^{h_m} S dz = \underbrace{\overline{w' \rho_s'}}_I + \underbrace{\int_0^{h_m} \frac{\partial \overline{\rho_s}}{\partial t} dz}_{II} + \underbrace{\int_0^{h_m} \bar{u} \frac{\partial \overline{\rho_s}}{\partial x} dz}_{III} + \underbrace{\int_0^{h_m} \bar{v} \frac{\partial \overline{\rho_s}}{\partial y} dz}_{IV} + \underbrace{\int_0^{h_m} \bar{w} \frac{\partial \overline{\rho_s}}{\partial z} dz}_{V} \quad (3.18)$$

where I represents the scalar source/sink term that corresponds to the net ecosystem exchange when the scalar is CO_2 and to the ecosystem evapotranspiration when the scalar is water vapor; II represents the eddy flux at height h_m (the flux which is directly measured by eddy covariance systems); III represents the storage of the scalar below the measurement height; IV and V represent the fluxes by horizontal and vertical advection. Under stationary atmospheric conditions and horizontal homogeneity, the last three terms of the right-hand side of equation (3.18) disappear and the eddy flux equals the source/sink term. However, in forest systems, these conditions are not always met and both storage and advection may be significant, especially at night [Aubinet *et al.* 2000]. Figure 3.10 summarize the air flow in the atmospheric boundary layer.

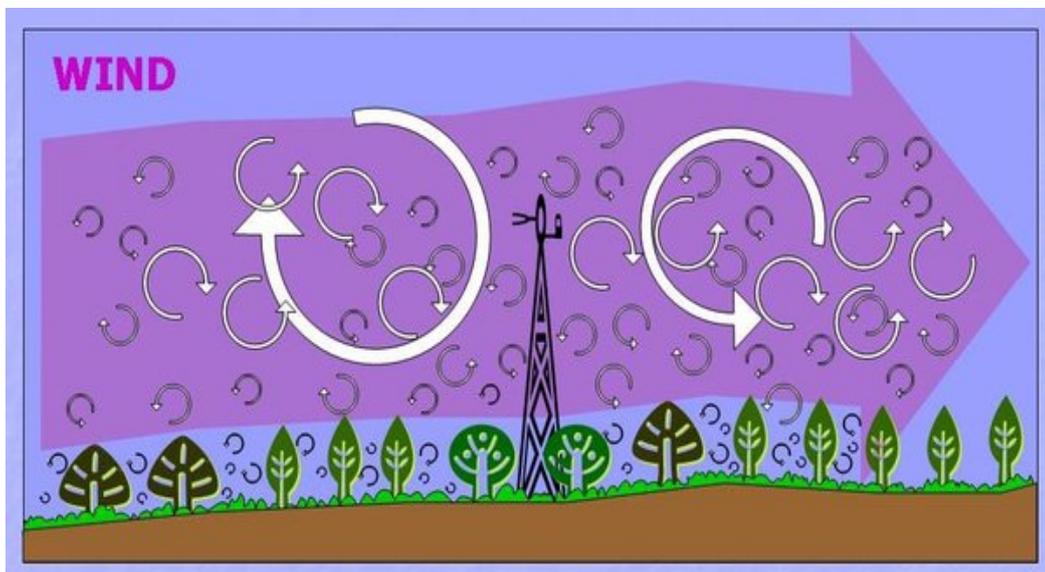


Figure 3.10. Air flow can be imagined as a horizontal flow of numerous rotating eddies (turbulent vortices of various size) with each eddy having 3D components. The situation looks chaotic, but vertical movement of the components can be measured from the tower.

The eddy covariance technique is used to cover as many forest types as possible, so we must consider sites that are not necessarily ideal for flux measurements, in particular, heterogeneous forests. One main problem in this case is to know to what extent the fluxes are representative of the real ecosystem flux. To know this, it is necessary to locate the flux sources and to establish the distribution of the frequencies at which they influence the flux measurements.

A first rule of thumb [Baldocchi *et al.*, 1988] was to consider that the boundary layer grows with an approximate angle of $1/100$, which means that the source is located within a distance equal to 100 times the measurement height above the effective surface. Anyway, this rule was too drastic

for eddy measurements, scalar flux source areas being smaller than scalar concentration areas by approximately one order of magnitude [Valentini, 2003].

Footprint models may help to refine this criterion. Footprint analysis was developed in order to estimate the source location. It is based on Lagrangian analysis and relates the time-averaged vertical flux of a quantity at the measurement point to its turbulent diffusion from sources located upwind from the measurement point. The extent to which an upwind source located a certain distance from the measurement point contributes to observed flux has been termed the source weight function or flux footprint. In footprint analysis, the relationship between the surface sources/sink and the measured flux is studied. Different analytical models were presented that differ in their complexity. They were based on the analytical diffusion theory. All these models showed that the source area size and upwind distance increased with the measurement height above the displacement height and with thermal stability, and decreased with surface roughness and thermal instability. By combining a simple footprint model with a site map, Aubinet *et al.* [2001] showed that, owing to the modifications of the flux source area with stability conditions, flux measurements made in heterogeneous forests could describe ecosystems that differ completely between night and day [Valentini, 2003].

The closure of the energy balance is a useful parameter to check the plausibility of data sets obtained at different sites. In this approach, the sum of turbulent heat fluxes is compared with the available energy flux (the net radiative flux density less the storage flux densities in the observed ecosystem, including soil, air, and biomass). According to the first law of thermodynamics, the two terms should balance, i.e., the regression of turbulent fluxes according to the available energy flux should have a slope of 1 and an intercept of zero.

Non-closure of the energy balance can be for different reasons. In a detailed analysis of the possible causes of energy balance non-closure, Wilson *et al.* [2002] identified five main primary error sources: (1) low and high frequency losses of turbulent fluxes, 2) neglected energy sinks, 3) footprint mismatch, 4) systematic instrument bias, 5) mean advection of heat and/or water vapour. Error 1 can be avoided by applying an appropriate correction. For error source 2, Wilson *et al.* [2002] cite energy storage in the soil above heat flux plates, latent heat losses below the heat flux plates, sensible and latent heat between the soil, and the measurement point. Besides, some processes like melting, freezing or heat conductance to cold rain are not considered in the budget. Error 3 results from the differences between the source areas of the different instruments measuring the energy fluxes [Schmid, 1997]. The source area of the radiometer is limited to a circle whose radius depends on sensor height and that is constant with time, while the source area of the eddy covariance system is approximated by an ellipse whose shape and position vary with wind and stability conditions. In these conditions, systematic difference between the

radiation and the turbulent fluxes may appear in heterogeneous sites. The systematic instrument bias (error 4) can affect the eddy covariance system and also the heat storage measurement systems and radiation instruments. Neglecting horizontal and vertical advection fluxes (error 5) could be one major reason for the lack of energy balance, in particular at night, during stable periods. As advection fluxes concern heat as well as water vapour and carbon dioxide, the lack of energy balance here could be an indicator of an underestimation of the carbon dioxide flux during this period.

The degree of closure of the energy balance may be used as a criterion to determine if the vertical turbulent fluxes do represent the total fluxes of a scalar or not. In particular, perturbation of the flux by the tower in some directions may be expressed by a lower degree of closure. On the other hand, the energy balance may serve as a tool for better analyzing night processes.

The eddy covariance do not provide directly measurement of both GPP and NPP; these last two variables, anyway, can be estimate for NEE measurement. On the other hand, the evapotranspiration can be directly computed by latent heat measurement.

In this work, daily values of ET and GPP, measured with EC technique [Reichstein *et al.*, 2005] in 10 different *CarboeuropelP* sites (Figure 3.11) [Valentini *et al.*, 2000], are compared with modeled data.

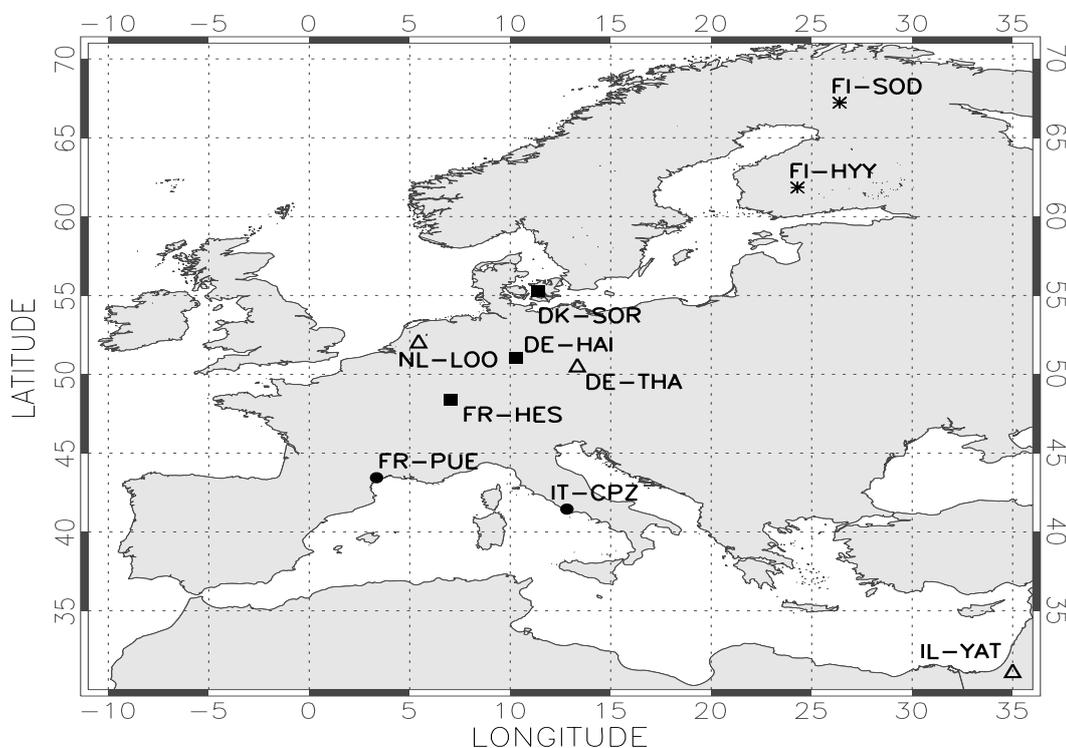


Figure 3.11. Sites chose to run data assimilation. Sites are divided by predominant PFT: squares indicate stations where predominant PFT is TBS, triangles indicate stations with predominant PFT composed by TNE, stars represent stations with BNE as the main PFT and finally circles represent sites where TBE is the main PFT.

The ten study sites were chosen in order to represent all the typical natural European forests. All the sites compute net ecosystem exchange of CO₂ using similar instrumentation and data processing techniques (*Baldocchi et al., 2001*)

These sites all have forested footprints with varying mixtures of deciduous and coniferous species. The sites are situated in different climate conditions: some basic features about the study sites are provided in *Table 3.7*.

Site Name	Site Code	Dominant species	Temp. (°C)	Prec. (mm/y)	Elevation (m)	Available Data
1 Hainich	DE-HAI	<i>Fagus sylvatica</i>	7	750	445	2000-2002
2 Sorø	DK-SOR	<i>Fagus sylvatica</i>	8.1	510	40	2000-2002
3 Hesse	FR-HES	<i>Fagus sylvatica</i>	9.2	885	300	2000-2002
4 Loobos	NL-LOO	<i>Pinus sylvestris</i>	9.8	786	52	2000-2002
5 Tharandt	DE-THA	<i>Picea abies</i>	7.5	820	380	2000-2002
6 Yatir	IL-YAT	<i>Pinus halepensis</i>	22	275	680	2001-2002
7 Hyytiälä	FI-HYY	<i>Pinus sylvestris</i>	3.5	640	170	2000-2002
8 Sodankylä	FI-SOD	<i>Pinus sylvestris</i>	-1	500	180	2001- 2002
9 Puechabon	FR-PUE	<i>Quercus ilex</i>	13.5	883	270	2000-2002
10 Castelporziano	IT-CPZ	<i>Quercus ilex</i>	15.3	550	3	2000-2002

Table 3.7. Main climatic features and dominant species of the study sites. Temperature refers to the annual mean.

A usual difficulty when applying inverse methods to geophysical problems is to properly define the uncertainties on the data, and to propagate them onto the parameters that are sought for. So far, the knowledge of the observational errors is difficult to retrieve for flux measurements [*Dore et al., 2003; Friend et al., 2006; Santaren et al., 2007*]. Error estimates across different sites range between ± 30 and ± 180 gC m² yr⁻¹, which are of the same order as the model-data difference [*Woodward and Lomas, 2004*], and may be even higher at some locations [*Kruijt et al., 2004*]. Sources of these errors include instrument calibration, data gaps, and atmospheric processes that reduce the correspondence between the measured flux and the ecosystem flux (e.g. advection or high atmospheric stability). In terms of cumulative (e.g. annual) NEE, random and fully systematic errors are generally of lesser concern than selective systematic errors, because the latter exist only for part of the daily CO₂ cycle when fluxes are either positive or negative [*Moncrieff et al., 1996; Friend et al., 2006*]. The model also has error in relation to the measured ecosystem through the assumption of natural (potential) vegetation with long-term equilibrium between fixation and respiration, system attributes unlikely to be true at most eddy flux sites.

Rigorous incorporation of these error terms into model-data comparisons remains a subject of on-going research [Hollinger and Richardson, 2005; Friend et al., 2006].

At the moment, there is no standard method to estimate the uncertainty associated to eddy data [Dore et al., 2003; Friend et al., 2006], so for each carbon flux less than $1 \text{ gC m}^{-2} \text{ d}^{-1}$ and water flux less than $1 \text{ mmH}_2\text{O m}^{-2} \text{ d}^{-1}$, we chose an error equal to 0.1, while fluxes greater than 1 have an uncertainty equal to 10% of their values. Covariances among observed fluxes are not considered.

Initially we ran the optimization using these values for the uncertainties of the observations; in this case we noted that the higher fluxes were underestimated at the end of the simulation. This was partially expected as a consequence of the analytical shape of $J_D(\mathbf{p})$ where the weights of squared deviations between predictions and observations are inversely dependent on the standard deviations associated to the latter. So, in order to avoid this systematical underestimation of higher fluxes, we used weights in $J_D(\mathbf{p})$ such that all the deviations had the same importance during the optimization process.

3.3 RESULTS: CHANGES IN MODEL PARAMETERS AND FLUXES AFTER A MONTE CARLO SAMPLING

3.3.1 Introduction

In this section we present the results of the model optimization. Initially, we compare prior and posterior fluxes, computed using the four optimization alternatives described above (*Post Best*, *Post Mean*, *Post 3 Best* and *Post 3 Mean*), with eddy fluxes in order to find out which of the four new parameterizations minimize the mismatch between simulations and observed data. Next, we describe the spatial and temporal variability of optimized parameters and we use the final-optimized parameters to run and validate the model for the year 2003. Finally, we analyze the effects of the new parameterization on the whole globe.

Generally data assimilation yields better fit between optimized variables and data, but could force other model variables to become unrealistic; for this purpose we used the final-optimized parameter values to check as well how changes GPP, ET and two variables not taken into account during optimization such as net ecosystem exchange and Total Ecosystem Respiration (TER).

3.3.2 Comparing different optimization strategies and performances

In order to compare fluxes and the results of optimization between the two different assimilation strategies, for each station we computed a single value for the whole period 2000-2002 as the mean of the three values obtained separately for the three years of assimilation (2000, 2001 and 2002); these are the optimized values computed within the optimization schemes *Post Mean* and *Post Best* valid for a single station.

When we compute the time-averages from single year optimization strategy we loose a lot of information we got employing MCMC method; for such reason it should be considered only be a first order approximation useful to compare parameters obtained from different assimilation schemes. The main difference between the two assimilation strategies is the time dependencies of parameters: in one case (3 years simultaneously) parameters are time-constant, while in the single year optimization, by definition they vary in time. So far, doesn't exist any model that make use of parameters varying in time, even if *Wang et al.* [2007] showed that some specific PFT parameters valid for deciduous forests vary seasonally as leaves develop.

Once we found an univocal value for each non specific PFT parameter for each station and for both assimilation schemes (*Post Mean* and *Post Best*), we ran the model, using the parameters we got from all different schemes, to compute water and carbon fluxes. To evaluate and compare model performances with respect to the new parameterizations we apply a *t*-test according to a slight modification of the method described by *Morales et al.* [2005]. This method is based on the daily deviations between observed (D) and simulated (M) values for the GPP and ET after optimization. We defined a *t*-Student variable as follows:

$$t = \frac{\bar{D} - \bar{M}}{\sqrt{\frac{s_D^2}{n} + \frac{s_M^2}{n}}} \quad (3.19)$$

where n is the number of paired values, a number generally lower than $365 \cdot 3$ (the number of days in the 3 years of assimilation) due to days with missing observations and s_D^2 and s_M^2 are the variances of, respectively, the observed data and the simulated values.

In the case of the null hypothesis (equal means) t has a zero expected value, so, for any given degrees of freedom and any confidence level, there is a critical t -value (t^*) beyond which we can refuse the null hypothesis. More precisely, if the t -value falls outside the confidence interval $(-t^*, t^*)$,

t^*), it can be concluded that the simulation shows a significant bias toward overestimation ($t < -t^*$) or underestimation ($t > t^*$) of the observations by the model.

We applied a two-tailed test with a significance level of 95%; in such case we always got a critical t -value equal to 1.963 for both modeled variables within all optimization schemes.

In *Figure 3.12* we plotted the t -values for prior GPP versus the t -values for prior ET for all 10 stations taken into account during optimization (*Table 3.7*). Since for all the stations the t -value falls outside the confidence interval ($-t^*$, t^*), it can be stated that the model systematically overestimates the evapotranspiration in DE-HAI, DK-SOR, FR-HES, FI-SOD and FR-PUE sites and underestimates it in all the other sites. As for GPP, only one sites (NL-LOO) falls inside the confidence interval ($t = -0.8$); the remaining sites show a GPP overestimation (IL-YAT, FI-HYY, FI-SOD and FR-PUE) or a significant underestimation (DE-HAI, DK-SOR, FR-HES, DE-THA, and IT-CPZ).

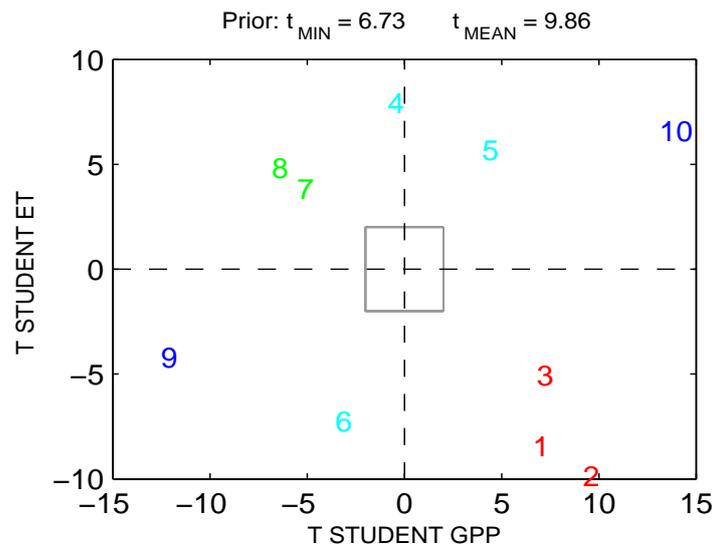


Figure 3.12. Scatter plots of two t -Student statistics describing the discrepancy between model and observations for the prior variables GPP and ET. Each point represents one of 10 different stations (*Table 3.7*), while colors indicate the PFT: azure indicate stations where predominant PFT is TNE, blue indicate stations with predominant PFT composed by TBE, green represent stations with BNE as the main PFT and finally red represent sites where TBS is the main PFT. The grey lines point out the thresholds (t^*).

In *Figure 3.13* we plotted the t -values for posterior GPP versus the t -values for posterior ET for all the assimilation schemes. Looking at the plot, we have a remarkable improvement in the results with respect to the prior case (*Figure 3.12*): in fact, almost all stations fall inside the confidence interval. From the statistical point of view, this means that there are no significant differences between the means of modeled and observed data.

Actually, looking at the four panels for evapotranspiration the t -values are almost all positive (they stay mostly above of x -axis); more precisely, for TBS there is a very good agreement between model and observed evapotranspiration: in fact all the t -values are overlapped or close to the zero axis in all plots. Two out of the three stations where Temperate Needleleaved Evergreen trees is the predominant PFT shows a good fit with measured ET, while the third (DE-THA) shows a slight underestimation in all assimilation schemes. For TBE, similarly, at one station (FR-PUE) the model matches well the observation in all cases, while at the other site (IT-CPZ) the t -values are far outliers in all four plots. For Boreal Needleleaved Trees the model, for evapotranspiration, shows a good agreement with observations in both sites and schemes.

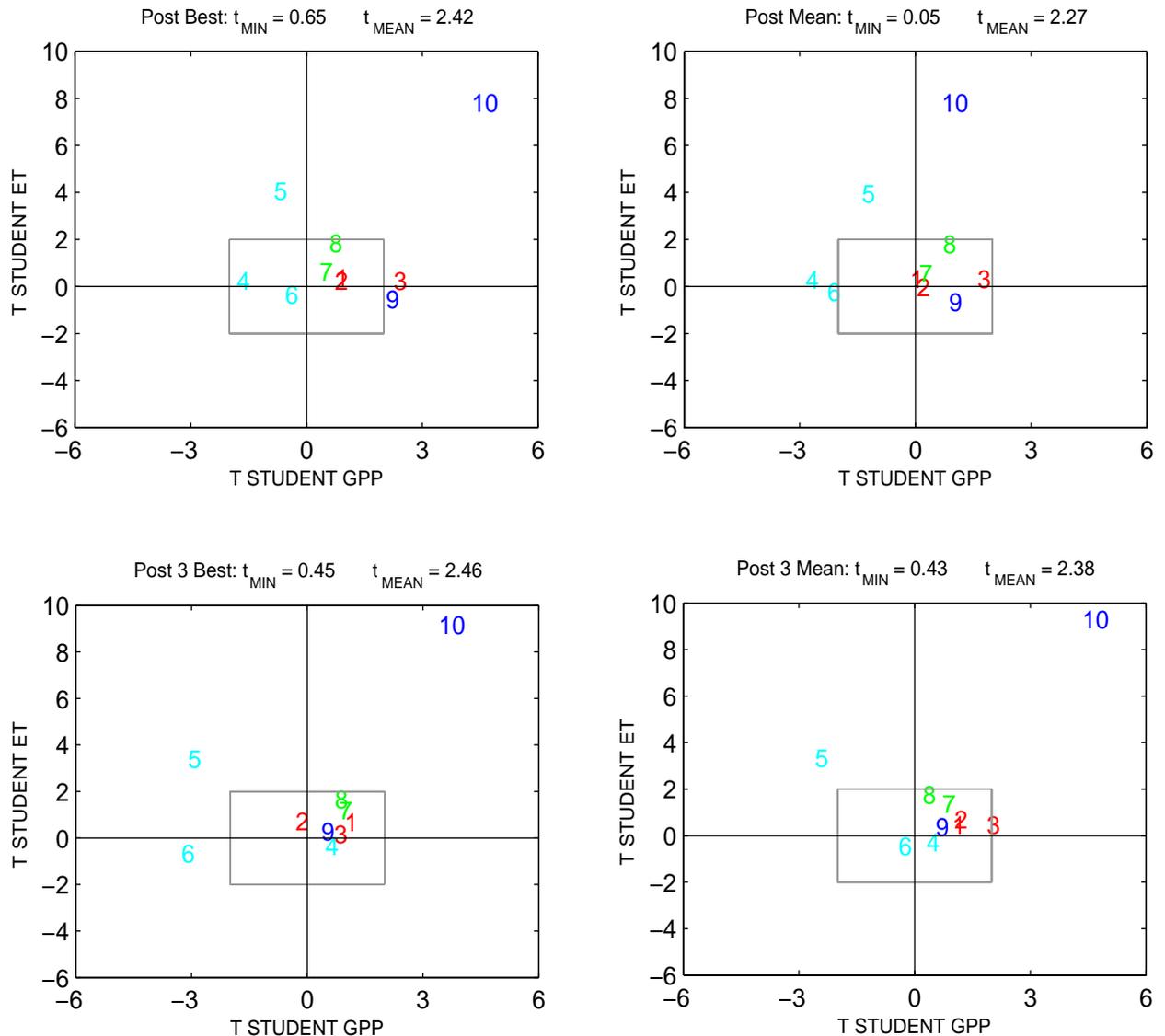


Figure 3.13. Scatter plots of two t -Student statistics describing the discrepancy between model and observation for the two optimized variables such as GPP and ET. Each plot refers to one of the four different schemes of optimization. Each point represents one of 10 different stations (Table 3.7). The grey lines represent the threshold (t^*).

About GPP, the t -values are distributed more uniformly with respect to zero but they have a tendency to be positive (they stay mostly to the right of y -axis). In particular, for TBS all the stations fall inside the confidence interval except FR-HES in the case of *Post Best* scheme. Also for the BNE all the stations fall inside the confidence. If we take into consideration TBE sites, at one station (FR-PUE) the model matches well the observation except in the case of *Post Best* scheme where the value is close to t^* ; at the other site (IT-CPZ) the t -values are far outliers in all plots except the *Post Mean* scheme. Finally, the performance of the model for TNE is less satisfactory because in all plots, except *Post Best* scheme, at least one station has a t -value beyond the critical value (t^*).

In order to identify the assimilation scheme with the minimum mismatch with respect to eddy fluxes, we computed, for each plot, the minimum (t_{MIN}) and the mean (t_{MEAN}) of the Cartesian distances of the ten stations from the origin of the axes. Looking at the values above the plots in *Figure 3.13* it appears that the best values for the mean and the minimum are in the second plot ($t_{MEAN} = 2.27$), so we point out the assimilation year by year and the mean of the second halves of random moves as the parameterization that give the best results for such data assimilation.

The results just obtained are confirmed by other two statistical methods of comparison between model and observations. For each station and assimilation scheme, we computed the squared correlation coefficients (R^2) and the Root Mean Square Error (RMSE) between model and data, both before and after optimization. In the case of *Post Mean* and *Post Best*, for both R^2 and RMSE, the relative increments have been computed year by year (starting from 2000 till 2002) and their average over the three years has been computed for each station (s) according to the following formulas:

$$\delta R^2(s) = \frac{1}{3} \sum_{i=2000}^{2002} \left(\frac{R^2_{Post} - R^2_{Prior}}{R^2_{Prior}} \right)_i \quad (3.20)$$

$$\delta RMSE(s) = -\frac{1}{3} \sum_{i=2000}^{2002} \left(\frac{RMSE_{Post} - RMSE_{Prior}}{RMSE_{Prior}} \right)_i \quad (3.21)$$

These increments are defined so that they are greater than zero when, running the model with optimized parameter values, the model performance improves with respect to the use of prior values. Finally we computed the averages of δR^2 and $\delta RMSE$ over all 10 stations.

Note, however, that we have four different possible posterior values for both R^2_{Post} and $RMSE_{Post}$, depending on which assimilation strategy has been used to perform the optimization;

so, the previous computations have been repeated in all four cases and the final averages have been plotted in *Figure 3.14* for comparison.

In the four panels of *Figure 3.14*, we present the results for each of the four variables GPP, ET, NEE, TER. We remind, however only two of those (GPP and ET) have been used to run optimization, while the other two (NEE and TER) are used as control variables in order to see if final parameter values produce an improvement in model performances also for variables not directly involved in the optimization process.

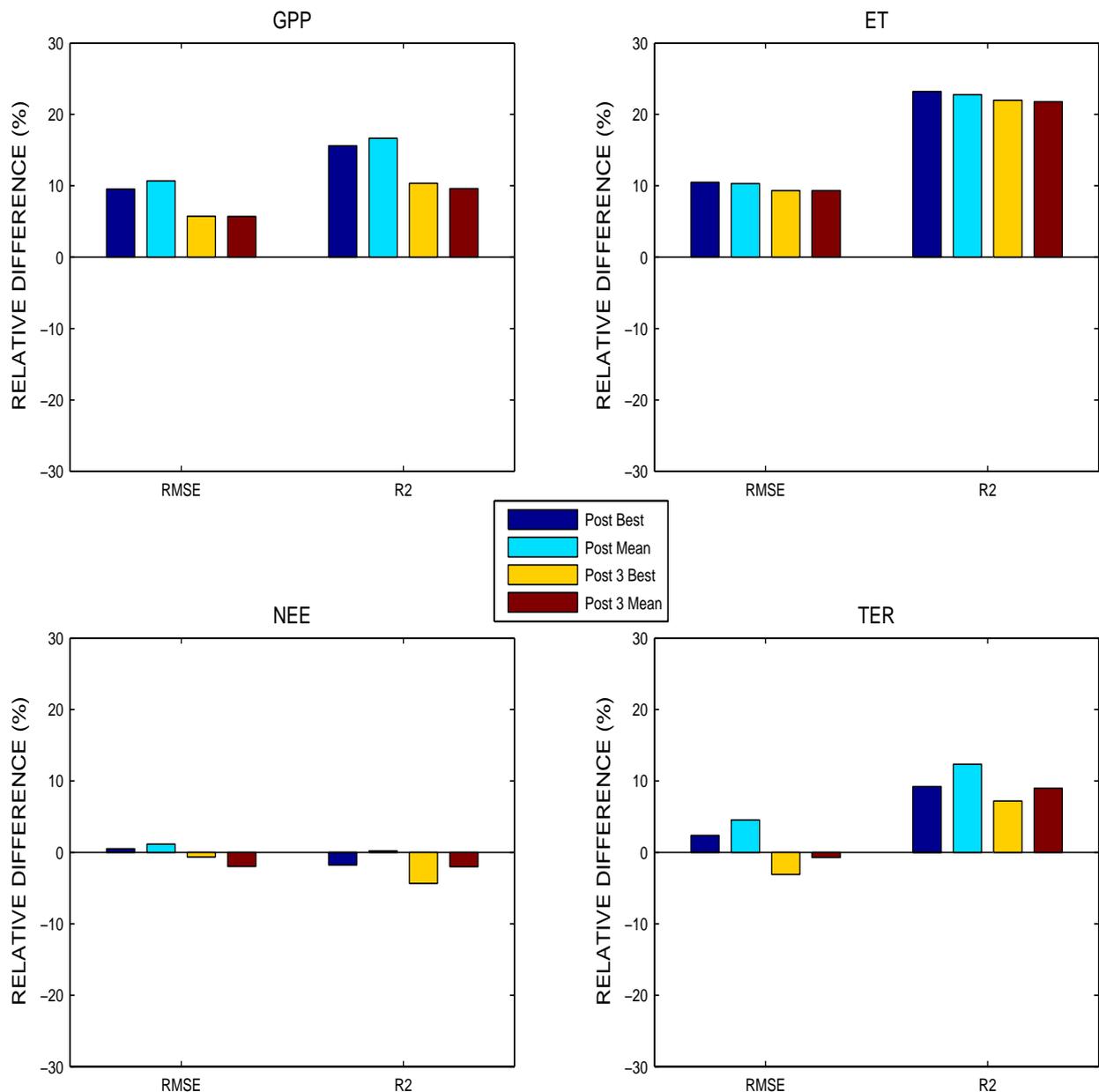


Figure 3.14. Relative increments between prior and posterior squared correlation coefficients (R^2) and root mean squared error (RMSE) for the different assimilation strategies and optimization schemes (different colors in the legend) for each of four variables (different panels).

Looking at the figure, the difference between the two upper panels with respect to the lower ones is remarkable: in the first case the optimization produces strong improvements in the ability of the model in simulating these variables, being all bars positive. In the other case, the negative bars indicate that the mismatch between data and model increases after the optimization.

But, as far as the *Post Mean* strategy (azure bars) is concerned, the model improves for all variables taken into account and for both statistical parameters (R^2 and RMSE). The other assimilation strategies have at least one statistical parameter lower than zero (or close to zero in the case of *Post Best* for NEE) in one panel and even when they shows positive values, these values are lower than *Post Mean* except for R^2 value for the evapotranspiration where all the bars are almost equal to each other.

Thus, also from this plot, we get a confirmation that, in the present work, the parameters computed within *Post Mean* assimilation scheme minimize the mismatch with respect to eddy fluxes. For this reason henceforth we compute final-optimized parameters values from this scheme of optimization.

3.3.3 Effect of different parameterizations on water and carbon fluxes

In this section, we analyze how water and carbon fluxes differ between the stations and the two assimilation schemes with the lower value of t_{MEAN} (*Post Mean* and *Post 3 Mean*, *Figure 3.13*).

As already stated above, in the case of year by year assimilation for each station we computed the posterior parameters as the mean over the three single years (2000, 2001 and 2002). In such way, we were able to run the model to check how the new parameterizations changed the fluxes and how the model with the new parameterization respond to the 2003 summer heat wave.

Respect to TBS, generally LPJ was able to simulates correctly the phase of the seasonal carbon cycle at all the three sites but the amplitude are somewhat underestimated using prior parameter (*Figure 3.15*, *Figure 3.16*, and *Figure 3.17*). As for evapotranspiration the model seems match the observation but with a slightly overestimation, as already shown above (*Figure 3.12*). For the years 2000-2002 both the data assimilation strategies works fine reducing the mismatch between eddy fluxes and simulations and only weak differences have been found between *Post Mean* and *Post 3 Mean* in the three stations. Considering 2003 heat wave, using both prior and posterior parameters the model matches quite good the observations: in particular, with prior parameters in Hainich (*Figure 3.15*) is overestimated the secondary summer peak of carbon assimilation of

deciduous trees even if it is well captured the anomaly, while with the posterior parameters the secondary peak is smoother than prior but the anomaly remains quite clear to be identified.

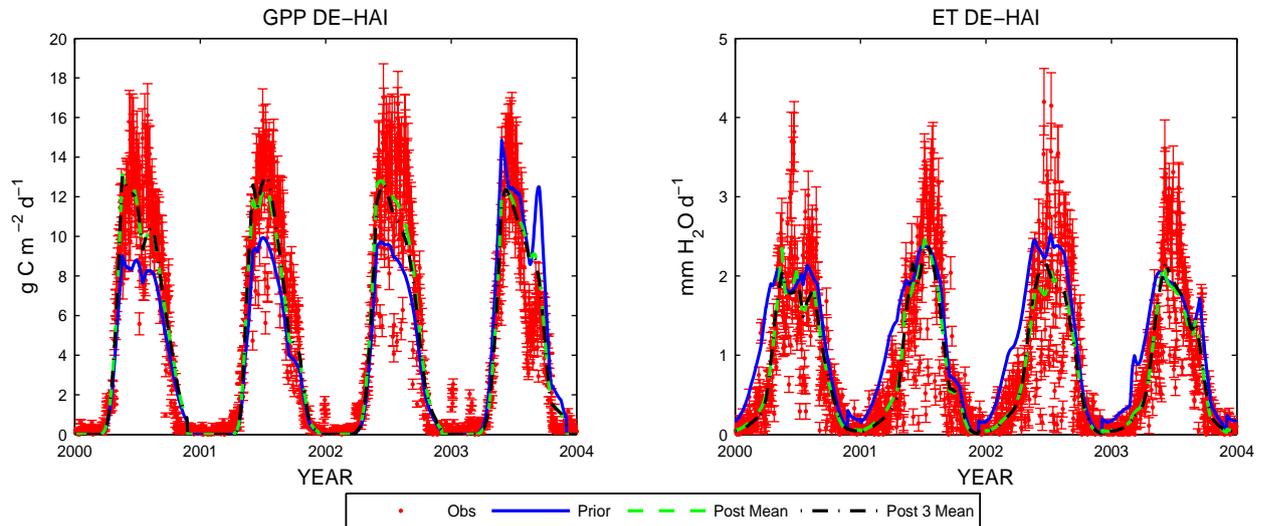


Figure 3.15. Observed vs. simulated daily fields of Gross Primary Production (GPP, $\text{gC m}^{-2} \text{d}^{-1}$) and evapotranspiration (ET, mm d^{-1}) for Hainich site .

Because of Sorø is the Northern TBS station, the heat wave anomaly is not so pronounced as other southern sites; anyway, here in 2003 with the prior parameters the GPP is deeply underestimated, while with the posterior parameters the simulations match the observed values (*Figure 3.16*).

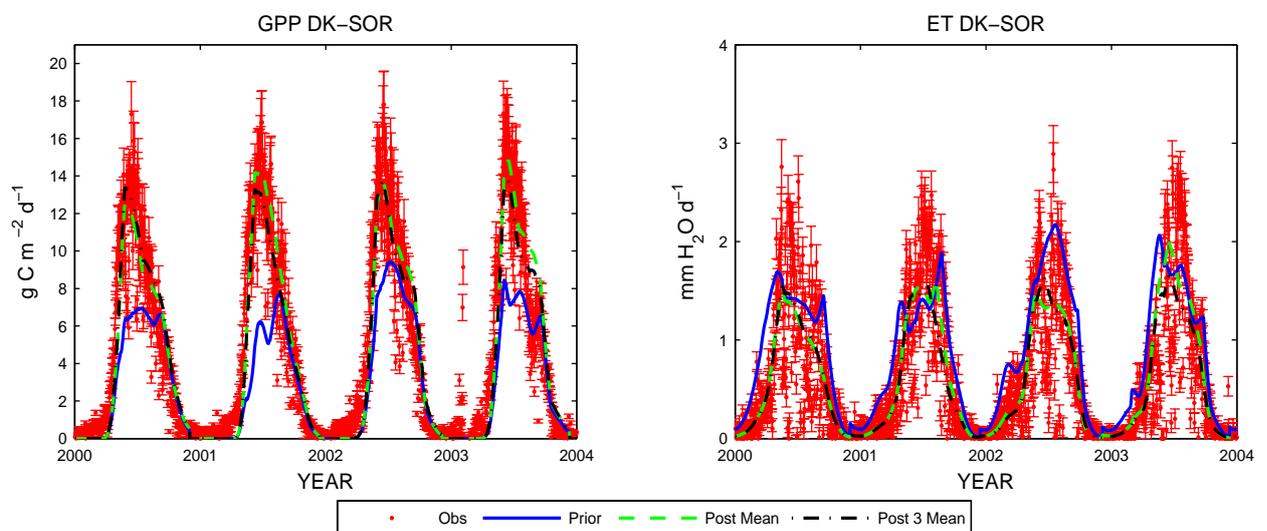


Figure 3.16. Observed vs. simulated daily fields of Gross Primary Production and evapotranspiration at Sorø site.

In Hesse (*Figure 3.17*) the results are less satisfactory than Heinich: in fact, with the prior parameters is underestimated the first carbon assimilation peak and consequently the anomaly is not pronounced, while using the posterior parameters generally the model matches the observation but also in this case the reduction of summer photosynthesis due to heat wave is weakly described and overestimated with both data assimilation strategies.

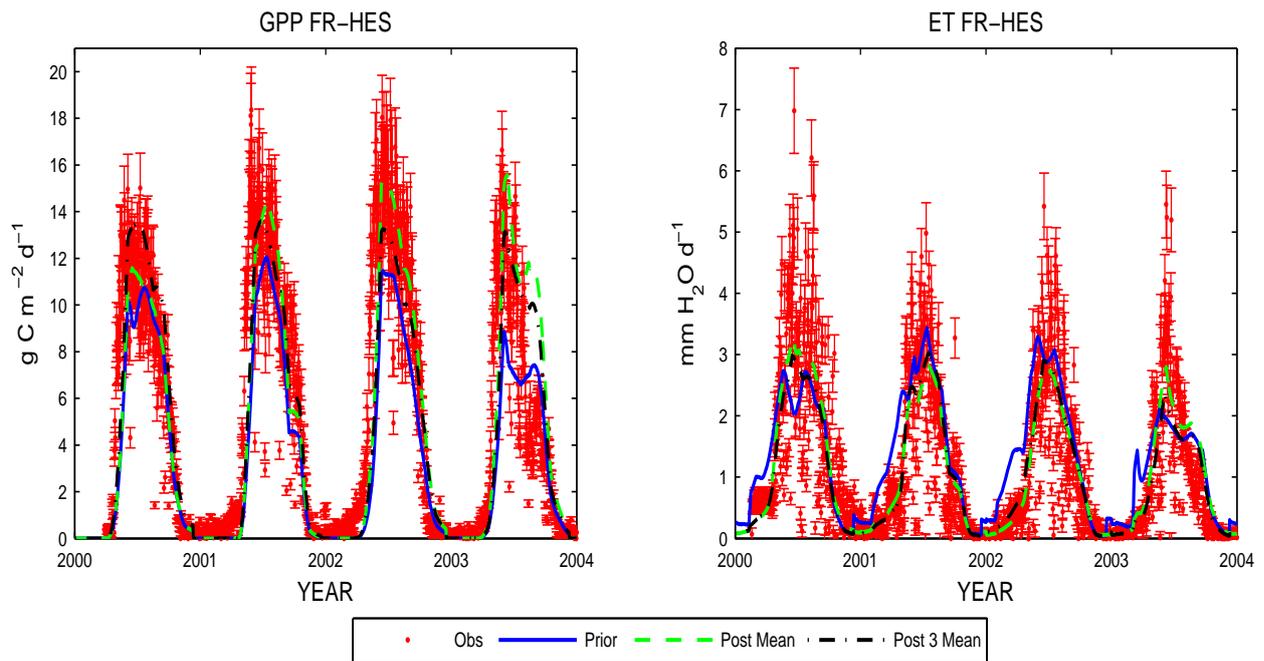


Figure 3.17. Observed vs. simulated daily fields of Gross Primary Production and evapotranspiration at Hesse site .

Looking at the results for TNE, LPJ (making use of prior parameters) was generally able to capture the observed seasonal cycle at all 3 stations (*Figure 3.18*, *Figure 3.19* and *Figure 3.20*) both for GPP and ET.

In Loobos only few differences occur between prior and posterior fluxes (*Figure 3.18*). As for year 2003 an abrupt decrease in GPP occurs during the summer for both assimilation strategies; this abrupt decrease in GPP is some days out of phase with respect to observations as well as the growing season starts some days before to the prior parameterization.

On the other side, the evapotranspiration matches with the eddy data (*Figure 3.18*) and only few differences have been found between prior and posterior fluxes. Considering years 2002 and 2003 the posterior fluxes show a different initial slope with respect the prior curve, but generally fluxes agree with measurement.

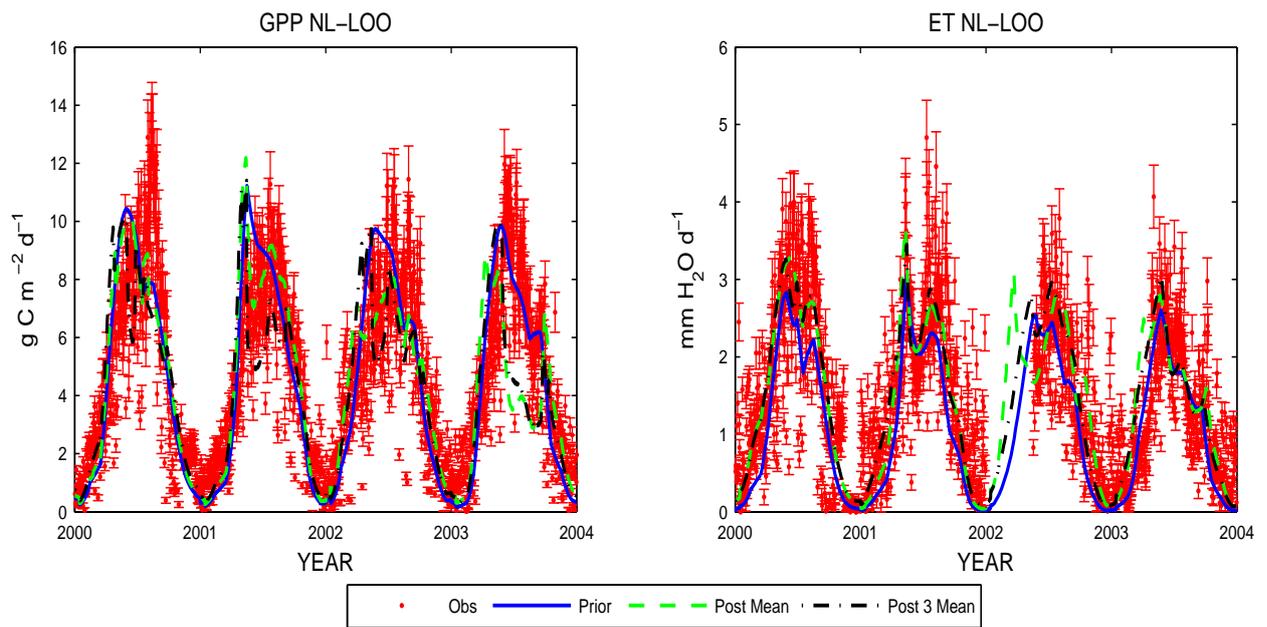


Figure 3.18. Observed vs. simulated daily fields of Gross Primary Production and evapotranspiration at Loobos site.

In Tharandt site few differences have been found between prior and posterior fluxes over the whole time period considered. In such stations LPJ simulates correctly the phase of the seasonal cycle and also the amplitude (Figure 3.19).

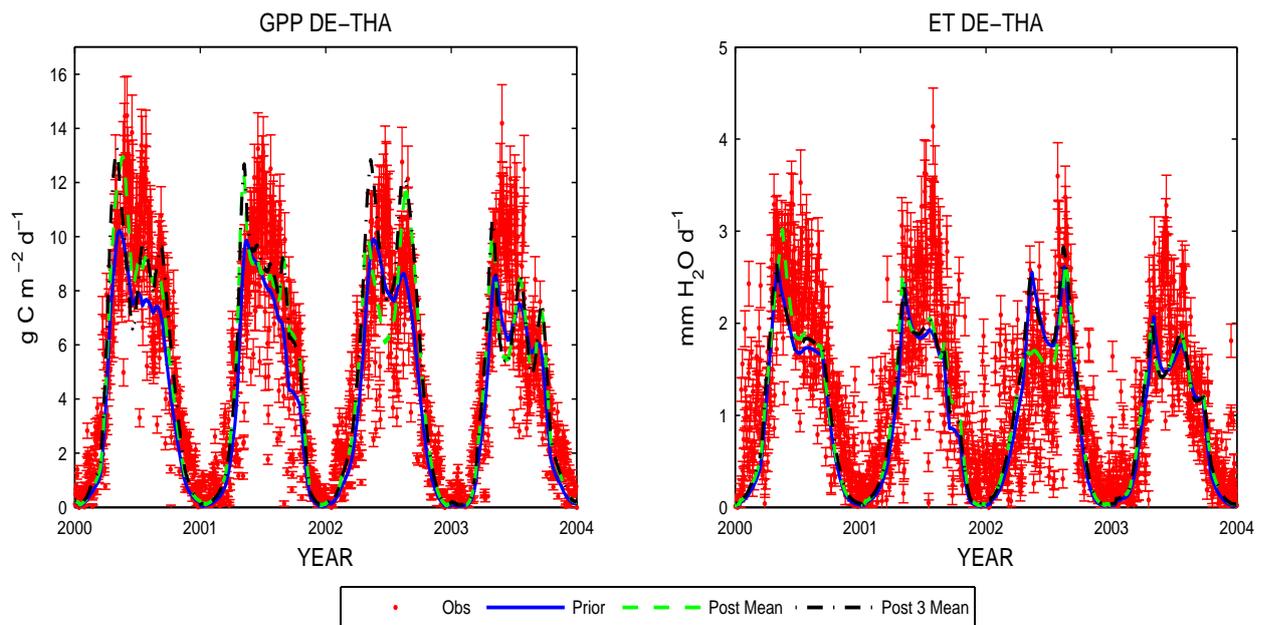


Figure 3.19. Observed vs. simulated daily fields of Gross Primary Production and evapotranspiration at Tharandt site.

In Yatir (*Figure 3.20*) we have the worst results: here, both GPP and ET are overestimated, as already shown in *Figure 3.12*. Moreover, considering the prior case, the carbon assimilation quickly declines to zero in the summer, while the observations do not show such behavior. We suppose this depends on the underestimation of the water availability during summer: in fact, unlike observations, simulated ET declines to zero and the trees respond to the drought closing the stomata and hence the photosynthesis is inhibited. After data assimilation, with both strategies the ET matches the observations and also the GPP improves.

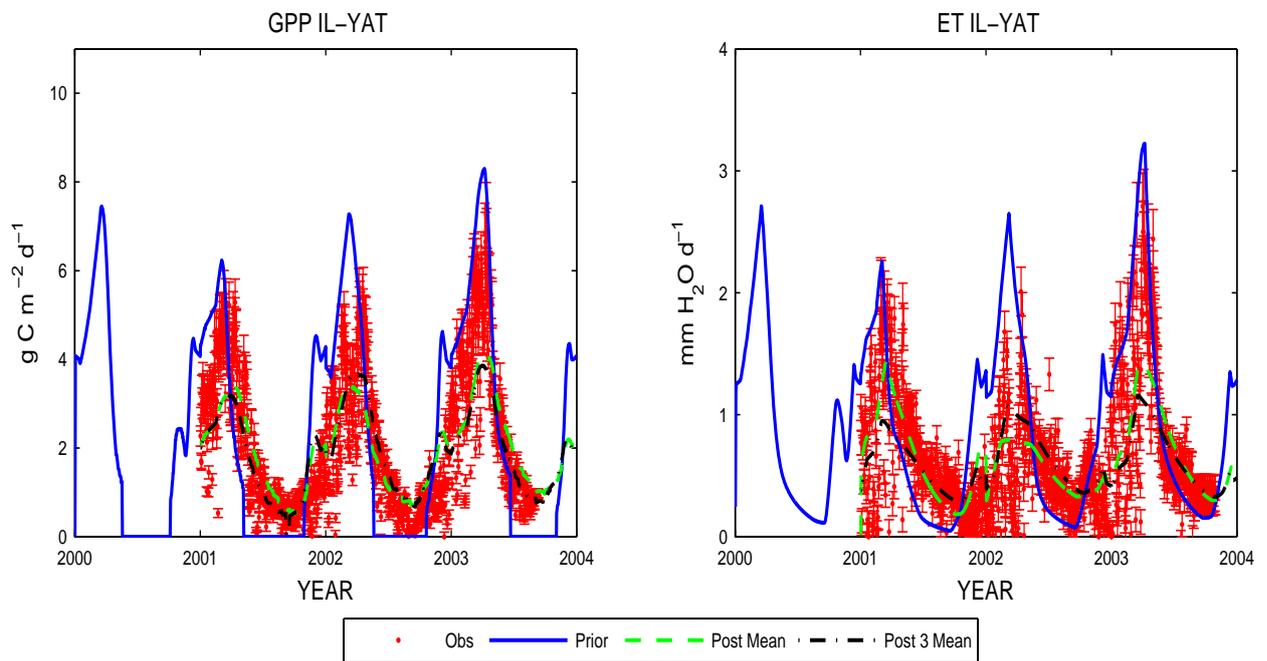


Figure 3.20. Observed vs. simulated daily fields of Gross Primary Production and evapotranspiration at Yatir site.

In both BNE sites the model using the prior parameters overestimates the GPP and underestimates the ET as already shown in *Figure 3.12*. In Hyytiälä (*Figure 3.21*) LPJ was able to capture the growing season but the amount of CO₂ fixed by photosynthesis is slightly greater than observed fluxes for the whole length of the simulation. The opposite behavior has been found for ET: for such variable, the model simulates correctly the phase of the seasonal cycle but significantly underestimates the water fluxes. Looking at the *Figure 3.21*, we can see that both assimilation strategies improve the model performances and no significant differences have been found between simulations and observation (see *Figure 3.13*) for the two variables taken into account. Moreover, the only differences between *Post Mean* and *Post 3 Mean* have been found only near the peaks of GPP and ET.

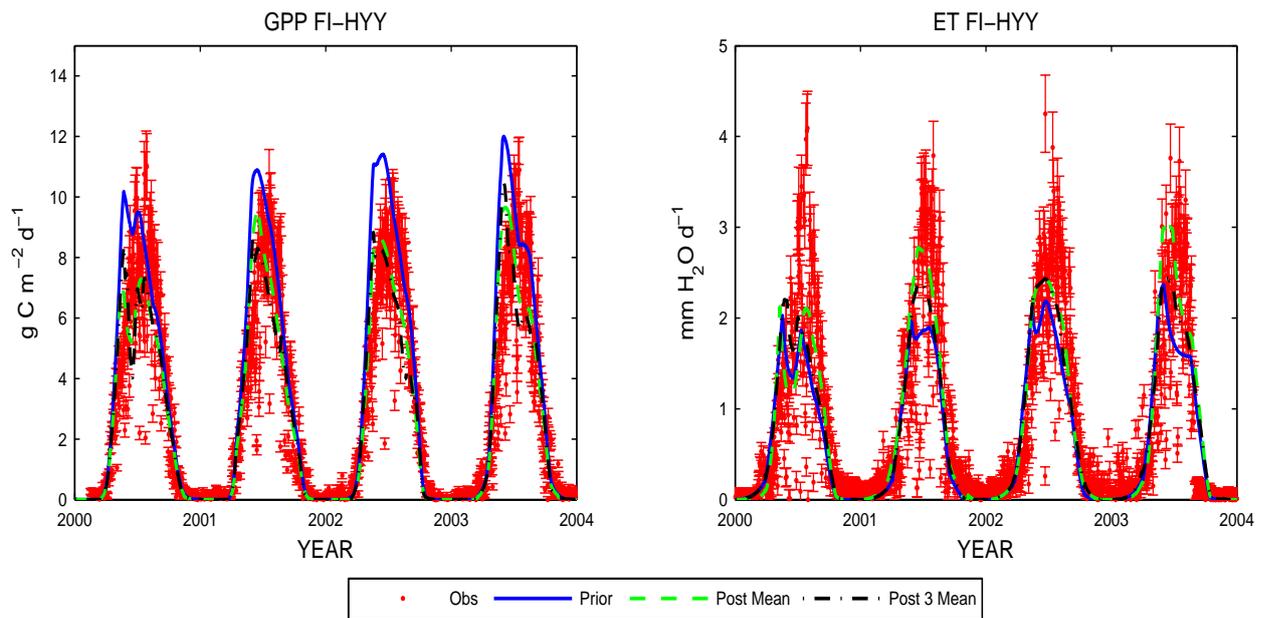


Figure 3.21. Observed vs. simulated daily fields of Gross Primary Production and evapotranspiration at Hyytiälä site.

In Sodankylä we have a similar behavior to Hyytiälä, but in this case the amplitude of GPP is widely overestimated, while for the ET the model seems simulate correctly the amplitude even if the t-test of *Figure 3.12* showed a slightly significant underestimation. This discrepancy depends on the uncertainty range that we associate to observations.

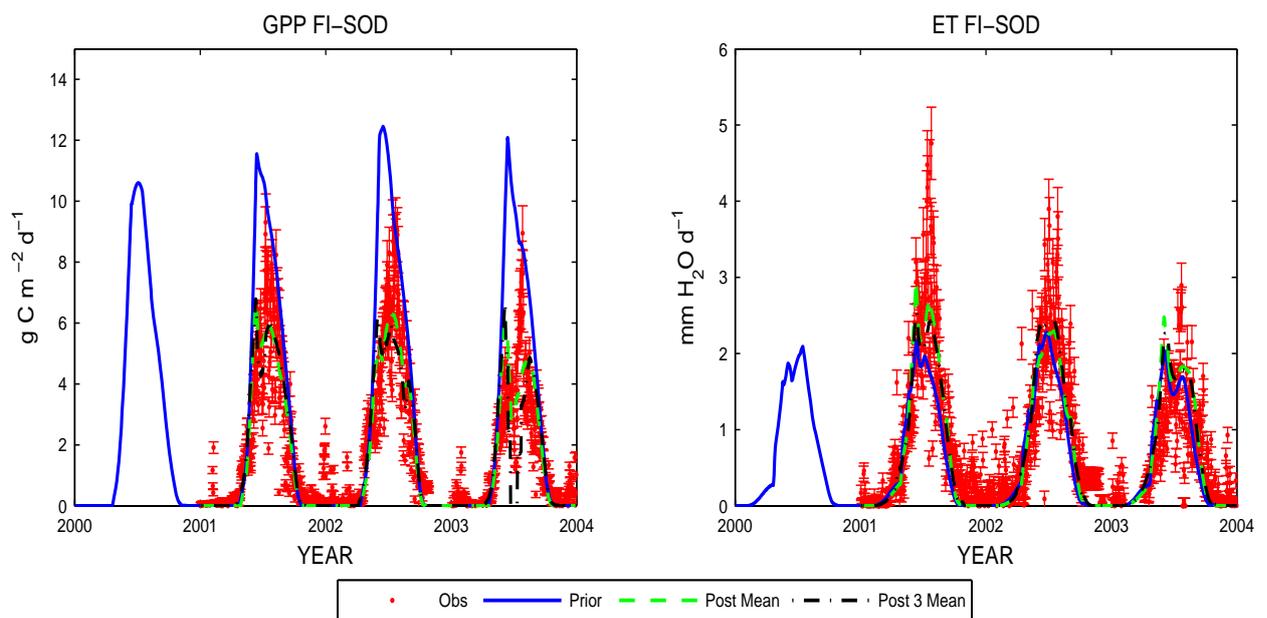


Figure 3.22. Observed vs. simulated daily fields of Gross Primary Production and evapotranspiration at Sodankylä site.

Finally, as for TBE sites the model with the prior parameters fits the phase of the GPP seasonal cycle in FR-PUE (*Figure 3.23*) but the amplitude is deeply overestimated (*Figure 3.12*). On the other side, there is a good agreement between simulated evapotranspiration and data, even if the t-test showed a slightly overestimation.

After data assimilation, the seasonal cycle of GPP remain practically unchanged, but the amplitude is noticeable decreased in both strategies; now the model matches correctly the observation and only few differences occur between *Post Mean* and *Post 3 Mean*.

As for ET with the prior parameters LPJ slightly overestimate water fluxes (*Figure 3.12*); after optimization is somewhere reduced the amplitude, while the phase remain unchanged.

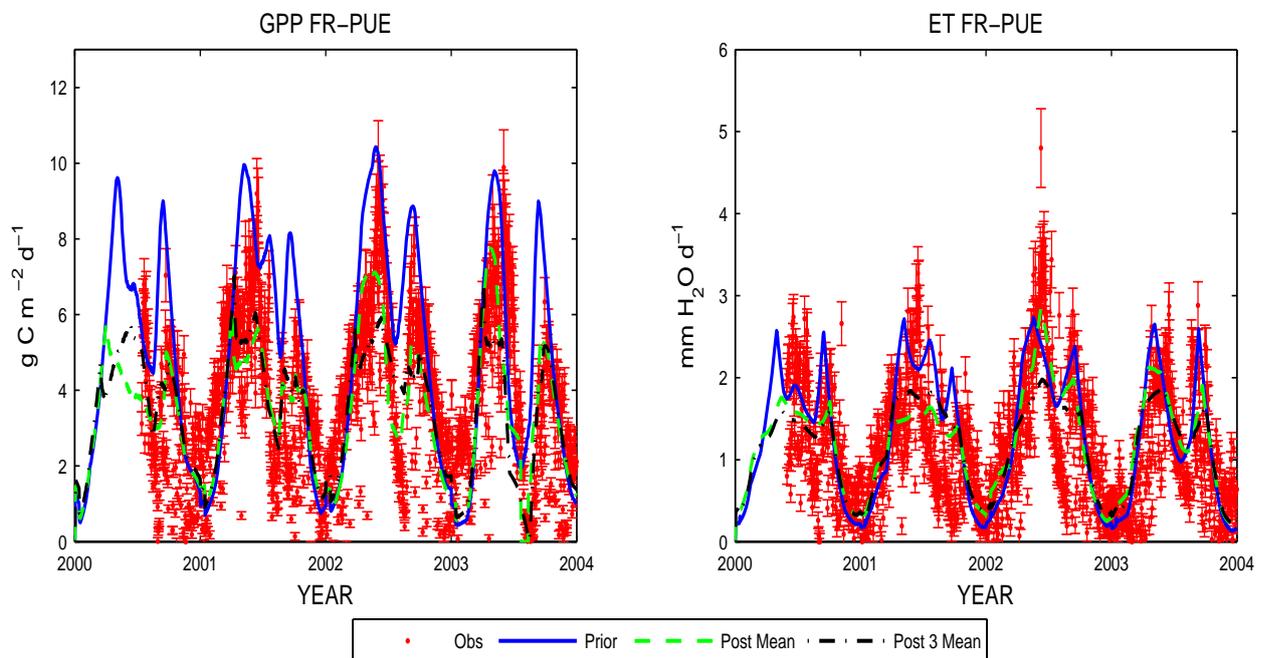


Figure 3.23. Observed vs. simulated daily fields of Gross Primary Production and evapotranspiration at Puechabon site.

Unlike Puechabon, in Castelporziano the GPP is underestimated in the prior case and the phase do not match the observations due to an abrupt decrease during the 2000 and 2001 summers. In such case, we got a similar behavior in Yatir, so also in this case we suppose that the reduction in photosynthesis might depends on the fast decrease of ET. After optimization, all the assimilation schemes adjust the ET reduction and the model simulates correctly both GPP seasonal cycle and amplitude during the assimilation period. Nevertheless during the 2003 summer we still have the same behavior; in fact, both posterior schemes and prior parameters show the same fast reduction in summer ET and this affect summer GPP as already stated above.

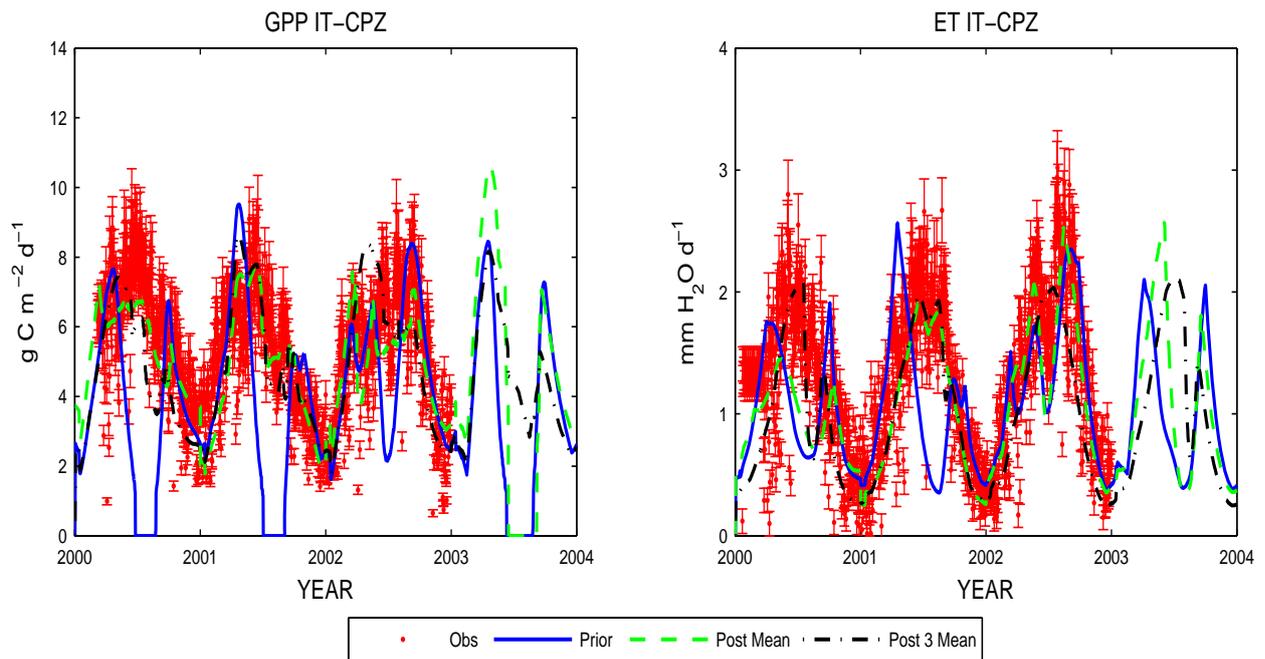


Figure 3.24. Observed vs. simulated daily fields of Gross Primary Production and evapotranspiration at Castelporziano site.

In the above figures LPJ have been benchmarked against eddy flux measurements at *CarboeuropelP* forest sites throughout Europe, and shown to produce acceptable fits to seasonal carbon and water fluxes. Performance tended to be poorer at Mediterranean sites, with representations of the mechanisms of response of plant physiology and allocation, and of microbial activity, to soil water deficits being identified as possible causes of model-data mismatches. This behaviour have been found also in other process-based models [Morales *et al.*, 2007]. Such behaviour reveal model deficiencies and data assimilation partially overcome such problem improving model-data fit.

3.3.4 Interstation and interannual variability

In order to analyze how the parameters vary in space and time, in *Figure 3.25* we plotted all non specific PFT posterior parameters we got from the two assimilation schemes that gave the lower number for t_{MEAN} (*Figure 3.13*). The label 00-02 refers to the *Post 3 Mean* optimization scheme, while the other three labels denote the values obtained for each single year optimization inside the *Post Mean* assimilation scheme.

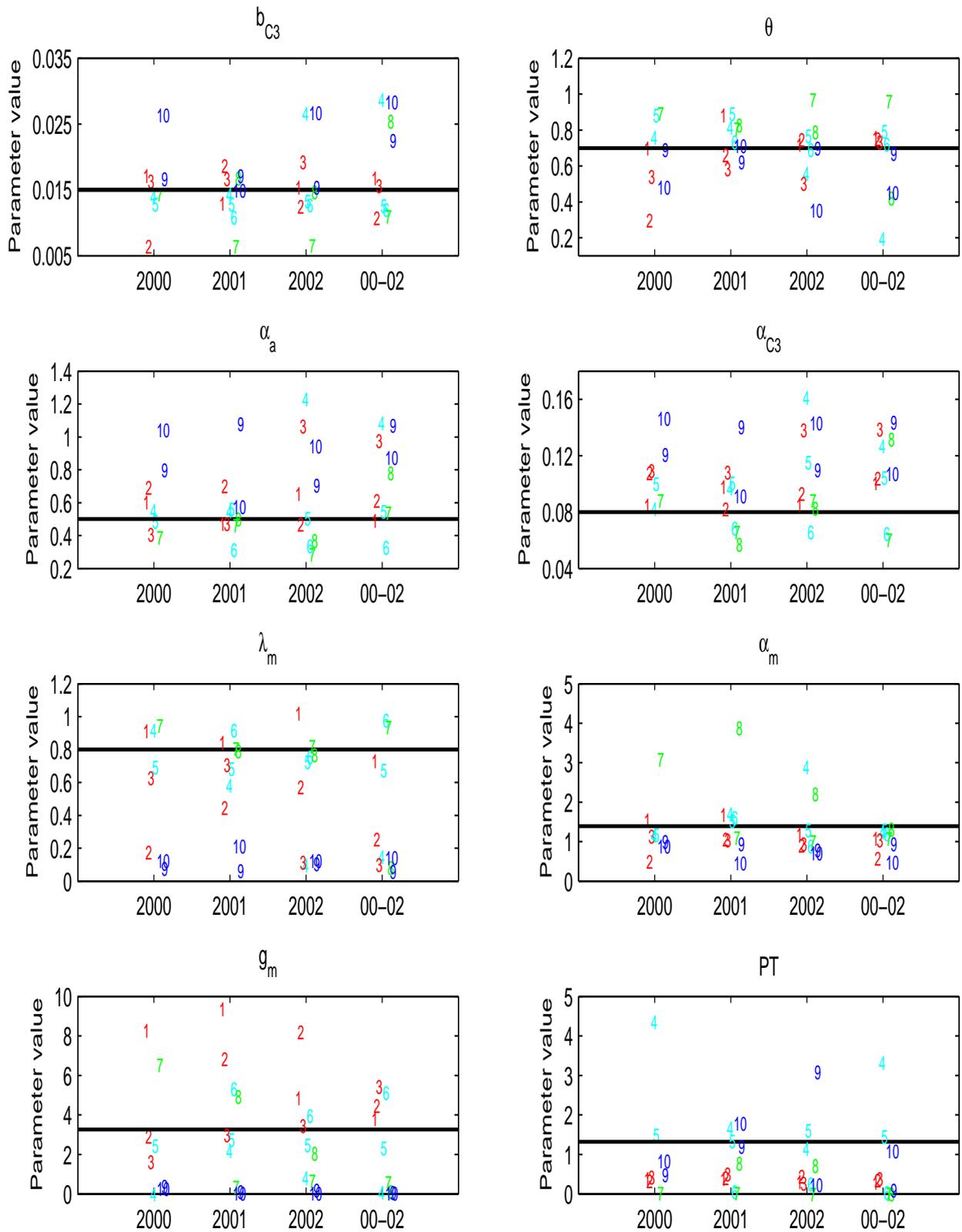


Figure 3.25. Spatial and temporal variation of optimized parameters for the two different assimilation schemes. Each number represents one of 10 different stations (Table 3.7), while colors indicate the PFT: azure indicate stations where predominant PFT is TNE, blue indicate stations with predominant PFT composed by TBE, green represent stations with BNE as the main PFT and finally red represent sites where TBS is the main PFT.

Looking at the different panels we infer that the spatial variability among different groups of stations (each group having stations with the same dominant PFT) seems, at first sight, to be remarkable for all the parameters, except for few parameters (bc_3 , α_m and PT in the case of TBS stations) where the values referring to stations with the same PFT appear to cluster together. Moreover, the interstation variability, for most of the eight panels, appears to be reduced in the *Post 3 Mean* case, especially when stations with the same PFT are considered. This might indicate that, when we need to assess a unique value for each non specific parameter, as the model requires, the *Post 3 Mean* optimization scheme seems to be preferable because this method decreases the spread among all the stations for almost all parameters.

Note that a high between PFTs variability could be an unfavourable result, since the eight parameters under examination should be, theoretically, independent of vegetation type; therefore, for each parameter, the between PFTs variability should be, possibly, comparable to the variability that we observe when we consider stations with the same PFT.

As for the interannual variability, for each given station, we point out that it is generally lower than the interstation variability but, however, it is not negligible and, moreover, it should be noted that for some parameters (α_m and g_m), the values obtained within *Post 3 Mean* optimization scheme often lies definitely out of the range containing the three values obtained for the single year optimization. This could be due to, essentially, two factors: first, the high non linearity of the model with respect to the all parameters under study and, second, the strong interference between different parameters during the optimization process because of high covariances between parameters implied by their joint PDF.

For most parameters, the statistical properties (mean, variance) computed over the ten stations for a given year generally shows a time-dependence: more precisely, when we move from year 2000 to 2002 parameters bc_3 and θ show a decrease of dispersion from 2000 to 2001 followed by an increase in the next year, while the mean do not present wide variations. For α_{C3} there is a marked increase in the dispersion around the mean, while the mean itself remain practically unchanged. A similar trend occurs also for α_a . In the PT panel, if we consider as outliers the two points far from the cluster for the years 2000 and 2002, we can assert that both the mean and the variance remain practically unchanged in time.

Moreover, it should also be noted that the spatial and temporal variability described above is not significant for almost all parameters and stations. MCMC is a probabilistic method and it does not provide a single best estimates but rather a PDF of posterior parameters; in other words to deliver the posterior estimates we should look both at the mean and the standard deviation. If we plot also the standard deviation associated to each parameter we found an overlapping of parameters both within and between PFTs and for different years. This mean that the interstation

and the interannual variability shown in *Figure 3.25* is mainly due to the posterior uncertainty associated to posterior parameters. This behavior, however, have not been found for all the parameters taken into account during optimization; for such reason, in some cases the above variability is relevant. One explanation for such variability may due to the eddy data. These represent, in fact, only fast processes (with an high influence of local conditions), while in LPJ the daily climatology is computed from monthly data using a linear interpolation. This lead to a smother evolution of the daily simulated variables such as GPP and ET and when we perform the optimization this may affect results. In other word the local-environmental conditions, captured with the eddy data and do not provided by the model, affect the results.

To understand how the optimal parameters vary in space between the two different assimilation schemes (*Post Mean* and *Post 3 Mean*), we computed a single value for the whole period 2000-2002, as already described above; this is the optimized value computed within the optimization scheme *Post Mean*. Finally, for each parameter, we computed the final value as the average over the ten stations; these averages are drawn as dashed lines in *Figure 3.26*. We also calculated the final parameters for *Post 3 Mean* scheme in the same way; these values are plotted in *Figure 3.26* as dash-dotted lines. The possibility of computing averages over all the stations, for the eight non-specific PFT parameters, is founded on the basic assumption that such parameters do not depend on the kind of vegetation present at each site. For the specific PFT parameters we can still compute the final-optimized parameter values by averaging over more than one station, but only over those sites with the same PFT. The necessity to adopt a univocal value for a given non specific PFT parameter in the model arises from the requirement that the model should be able to simulate a given process by using the minimum possible number of tunable parameters and to make predictions for all the regions of the world having the maximum fitting with the measurement. In fact, process-based models have been developed to reproduce broad general patterns of water and carbon exchange and vegetation dynamics at different scale for the whole globe using some parameters valid for all the biomes. The validity of this basic assumption is partially confirmed by the results showed in *Figure 3.26*: this shows how there is a low variability among the ten stations for both assimilation schemes and for almost all the eight model parameters. In particular, some parameters, such as b_{C3} , θ , α_a , α_{C3} and α_m , have a modest interstation variability, while the others show a larger interstation variability; however, the interstation variability appears lower if we consider only one of the two assimilation schemes.

As matter of fact, for some parameters (λ_m , g_m and PT) and stations, the two assimilation schemes yield remarkable differences in both senses in the values computed during the two optimizations. Once again, we suppose this is due to the high values of the covariances among parameters, as explained above.

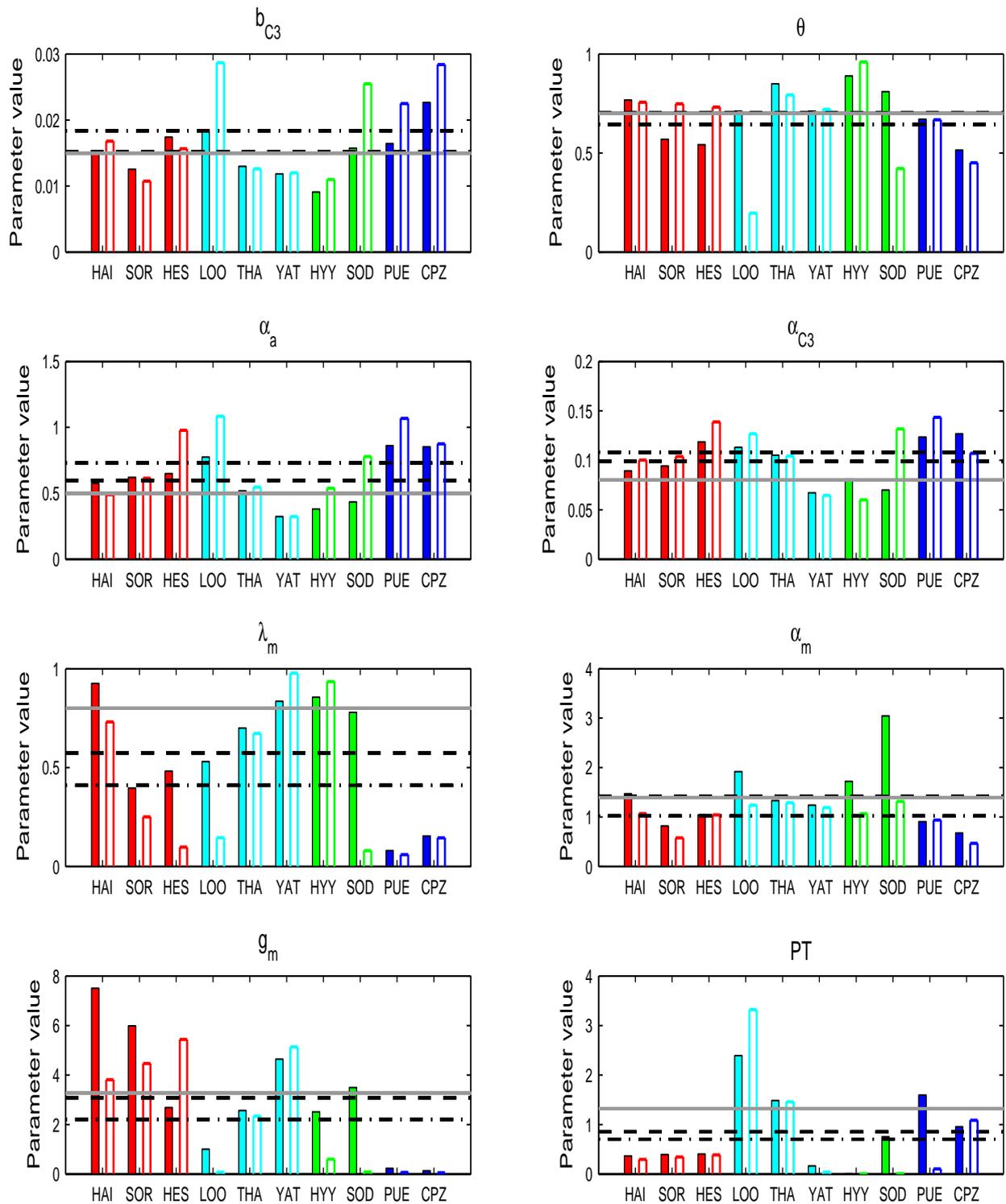


Figure 3.26. Each of the eight panels shows, for a given parameter, the variability among stations of the optimized parameters for both assimilation schemes. For each site we have 2 bars: the filled bars refer to Post Mean, the latter to Post 3 Mean. Different bar colors refers to different vegetation types: red indicates stations where predominant PFT is TBS, green represents sites where BNE trees is the main PFT, blue represents stations with TBE trees as main PFT and finally azure indicates stations with predominant PFT composed by TNE. The solid horizontal lines indicate the prior values, the dash-dot lines show the mean over all the stations for Post 3 Mean, while the dashed lines represent the mean over all the stations for Post Mean.

The final-optimized values, indicated as horizontal lines, are very close to each other and to the grey solid line for almost all the parameters taken into account during the inversion: in the case of parameters that show a large interstation variability (e.g. g_m and PT) this depends on the fact that the above differences occur randomly in both ways for each parameter and generally tend to compensate with each other, while where a low variability occurs the result is expected. Also, the conservative value of prior standard deviation imposed to the parameters (see *Table 3.9*) during optimization tends to produce a set of posterior parameters not so far from the prior ones.

Figure 3.26 shows one of the most important results of this thesis. Here we pointed out how there is a modest and, in such cases, significant interstation variability in both assimilation schemes but, when we compute the final averages over all sites, these are again close to priors values. In other word, this mean that the model and parameterization is not designed to deliver good estimates at single stand sites (local scale) but on larger regions or on the whole globe.

Obviously, at a site the local-climate influence is very high and the model can only cope with it by changing posterior parameters in order to deliver the best estimates of water and carbon fluxes. On the other hand, on larger regions the prior parameters were quite capable producing reasonable estimates and the average over several sites of posterior parameters confirms how the model is designed to simulate fluxes in a wide amount of different biomes.

The above results, nevertheless, give just a qualitative sketch of the spatial variability of the different parameters. At a first sight, it might appear that such variability depends on the dominant vegetation and, hence, the parameters could not be considered PFT-independent, while for a more quantitative assessment we perform a standard statistical analysis.

So, in order to asses whether the above spatial parameters variability is significant and also to give a more quantitative interpretations of the results, we performed an ANOVA test grouping the ten stations in four groups, each of these having the same dominant PFT; results are shown in *Table 3.8*. The null hypothesis for each parameter is that the four means computed within every group are not different and therefore they match with the overall mean computed over the ten stations; it is common to reject the null hypothesis if the p-value is less than a threshold given by the chosen confidence level; we adopt a 95% confidence level so the corresponding p-value is 0.05.

Because the p-values in *Table 3.8* are greater than 0.05 for all the parameters, except λ_m in the *Post Mean* assimilation scheme which is slightly less than 0.05, we accept for all parameters and both assimilation strategies the null hypothesis: in other words, the spatial parameters variability shown in *Figure 3.26* is not statistically significant at 95% confidence level that is, contrarily to what might appear from the figure, it can be stated that each parameter has the same mean, independently of the kind of vegetation. Since these parameters, examined with the ANOVA, are

by definition PFT-independent, this result is not unexpected and confirms how the parameters of the Farquhar model are independent from the kind of vegetation. The only parameters for which the p-value are close to the threshold are λ_m and g_m in the *Post Mean* and *Post 3 Mean* assimilation strategies, respectively. In these cases, the corresponding F value, that represents the ratio of the between variance to the within variance, is remarkably higher indicating that the between groups variability is five times higher than the within groups variabilities.

Par		Variability	MS	F	p	Variability	MS	F	p
b_{C3}	Between PFT	6.0E-05	2.0E-05	1.50	0.32	1.5E-04	5.0E-05	0.94	0.48
	Within PFT	8.0E-05	1.3E-05			3.2E-04	5.3E-05		
θ	Between PFT	9.3E-02	3.1E-02	3.22	0.10	6.6E-02	2.2E-02	0.35	0.79
	Within PFT	5.8E-02	9.7E-03			3.8E-01	6.4E-02		
α_a	Between PFT	2.2E-01	7.3E-02	4.11	0.07	1.5E-01	5.0E-02	0.62	0.63
	Within PFT	1.1E-01	1.8E-02			4.8E-01	8.0E-02		
α_{C3}	Between PFT	2.6E-03	8.5E-04	2.90	0.12	1.3E-03	4.3E-04	0.42	0.75
	Within PFT	1.8E-03	2.9E-04			6.1E-03	1.0E-03		
λ_m	Between PFT	5.8E-01	1.9E-01	5.42	0.04	3.2E-01	1.1E-01	0.69	0.59
	Within PFT	2.1E-01	3.6E-02			9.4E-01	1.6E-01		
α_m	Between PFT	3.0E+00	9.8E-01	4.25	0.06	4.6E-01	1.5E-01	3.06	0.11
	Within PFT	1.4E+00	2.3E-01			3.0E-01	5.0E-02		
g_m	Between PFT	3.3E+01	1.1E+01	3.45	0.09	3.3E+01	1.1E+01	4.66	0.05
	Within PFT	1.9E+01	3.2E+00			1.4E+01	2.4E+00		
PT		2.2E+00	7.3E-01	1.46	0.32			1.30	0.36
	Between PFT	3.0E+00	5.0E-01			3.8E+00	1.3E+00		
	Within PFT					5.9E+00	9.8E-01		

Table 3.8. Synthetic prospect of ANOVA results for the eight non-specific parameters. The column labelled “variability” shows the sum-of-squares (SS); the first row refers to variability between the four different groups (each group having stations with the same dominant PFT), while the latter refers to the mean of the variabilities each computed within the same group. The degrees of freedom associated to both the between and the within variabilities are 3 and 6, respectively, for all the parameters under examination. The MS column is the ratio between the corresponding SS and its degrees of freedom, while F represents the ratio between the two MS values. Finally, the last column indicates the probability to obtain the given F value or higher under the null hypothesis. Light gray shaded area refers to Post Mean assimilation strategy, while dark grey refers to Post 3 Mean.

3.3.5 Final-optimized parameters and uncertainty

Figure 3.27 shows normalized prior and final-optimized parameter values for the *Post Mean* assimilation scheme, with their prior and posterior standard deviations; the normalized parameter values are defined as the ratio between the dimensional parameter value and the respective prior value.

The figure shows that some optimized parameters values (b_{C3} , θ , α_m) are very close to their corresponding prior values, some others (λ_m and PT) are remarkably lower, while α_a and a_{C3} have posterior values slightly higher than prior ones.

As for the uncertainty associated to each parameter, there is a reduction of the standard deviation with respect to the assumed prior values for all parameters, but for g_m the posterior standard deviation is remarkably increased.

Note that this result still may depend on the prior uncertainty we associated to the prior parameters [Knorr and Kattge, 2005]; in absence of better knowledge, it was only estimated in a simple and preliminary way for this study. Also, assimilating more years of data or performing data assimilation on more sites using different values for prior uncertainty would lead to stronger constraints of model parameters and fluxes, which would lead to even smaller uncertainty of the posterior parameters.

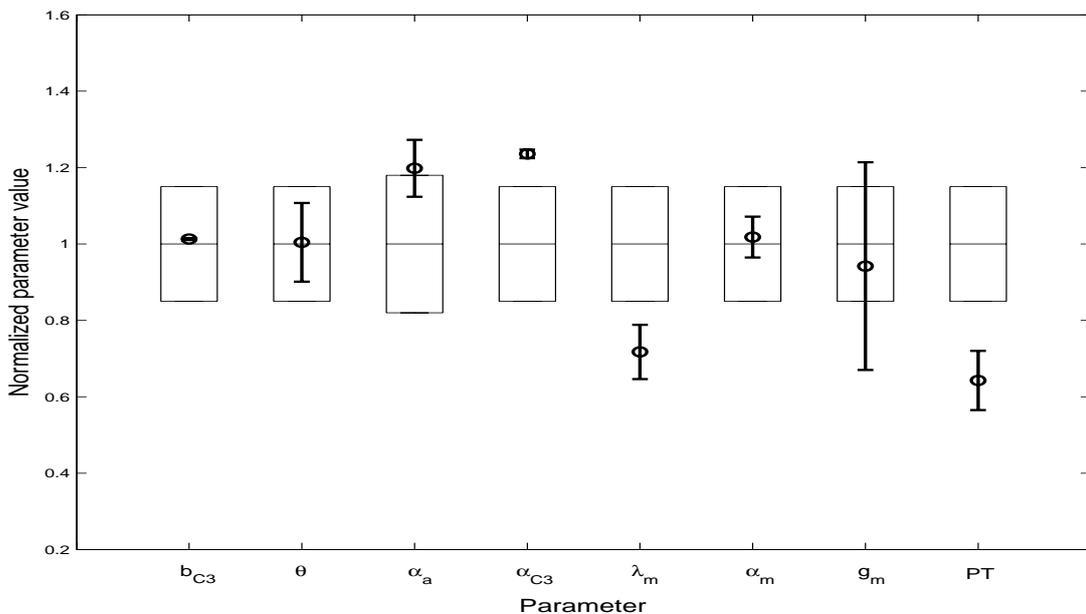


Figure 3.27. Prior and posterior parameter values and uncertainties for the normalized parameters. The boxes show the prior parameter values and their associated standard deviations; circle indicates optimized-posterior values and the error bars denote the standard deviation associated to the posterior parameters.

Table 3.9 summarizes the prior and posterior parameter values, along with their standard deviations.

Name	Prior value	Range	Post value	Prior SD	Post SD
b_{C3}	0.015	0.013-0.017	0.0152	15%	0.3%
θ	0.7	0.595-0.805	0.703	15%	10%
α_a	0.5	0.41-0.59	0.599	18%	7%
α_{C3}	0.08	0.068-0.092	0.099	15%	1%
λ_m	0.8	0.68-0.92	0.574	15%	7%
α_m	1.391	1.182-1.6	1.41	15%	5%
g_m	3.26	2.77-3.75	3.07	15%	27%
PT	1.32	1.12-1.52	0.848	15%	8%

Table 3.9. Prior and posterior values for model parameters and their standard deviations.

PFT	Name	Prior value	Range	Post value	Prior SD	Post SD
<u>TNE</u>	z_1	0.6	0.51-0.69	0.588	15%	3%
	E_{max}	5.0	4.25-5.75	5.96	15%	25%
	g_{min}	0.3	0.255-0.345	0.271	15%	3%
<u>TBE</u>	z_1	0.7	0.595-0.805	0.345	15%	1%
	E_{max}	5.0	4.25-5.75	6.62	15%	15%
	g_{min}	0.5	0.425-0.575	0.154	15%	4%
<u>TBS</u>	z_1	0.7	0.595-0.805	0.714	15%	4%
	E_{max}	5.0	4.25-5.75	6.08	15%	36%
	g_{min}	0.5	0.425-0.575	0.46	15%	9%
<u>BNE</u>	z_1	0.9	0.765-0.95	0.63	15%	2%
	E_{max}	5.0	4.25-5.75	6.75	15%	41%
	g_{min}	0.3	0.255-0.345	0.356	15%	5%
<u>GRASS</u>	g_{min}	0.5	0.425-0.575	0.504	15%	8%

Table 3.10. Prior and posterior values for specific PFT parameters and their standard deviation.

Table 3.10 shows how posterior values and uncertainty vary for the specific PFT parameters. The value of E_{max} has a very high posterior standard deviation for all PFTs except for TBE. From Table 3.9 and Table 3.10 we can see that some posterior values fall outside the apriori range (column 3 and 4, respectively) prescribed by means of standard deviation. It means that the

Bayesian term $J_0(\mathbf{p})$, depending on the prior probability, is negligible with respect to the term expressing the misfit between the model and observations $J_D(\mathbf{p})$, so that prior values are practically ignored during the optimizations.

Besides the standard deviation associated to the mean value of a given parameter, the MCMC method allow to compute also the correlations between any couple of parameters taken into account during the optimization. The correlation matrix helps to identify groups of parameters that tend to be constrained together [Knorr and Kattge, 2005].

Therefore, for any single station we can build a correlation matrix. The ten matrices thus obtained are rather different between them, so that for a given pair of parameters some stations show a significant positive correlation while other stations have a significant anti-correlation.

We show in the right panel of *Figure 3.28* the elementwise mean of these ten correlation matrices and in the left panel the maximum values of both correlations and anti-correlations found searching over all the ten stations.

As for the right panel of *Figure 3.28*, first of all we underline that all the filled boxes indicate highly significant correlations (even when the absolute correlation value is low) because of the huge amount (many hundreds of thousands) of data available in the Markov chain after the burn-in time [Knorr and Kattge, 2005; Braswell et al., 2005].

Usually, correlation values (r) above (below) 0.8 (-0.8) indicate pairs or groups of parameters that tend to be strongly constrained together [Quinn and Keough, 2002]. Nevertheless, in the present work, we believe that also correlation values as low as 0.2 point out pairs of parameters with a reliable, though modest, degree of mutual interrelation.

By comparing the two panels of *Figure 3.28* we underline that, although local conditions at each site may affect the correlation of a given pair of parameters in a way that might produce a notable spread of correlation values of both signs, after averaging over stations we are still able to retrieve some remarkable correlations among different groups of parameters.

In particular, α_{C3} and α_a are strongly anticorrelated ($r \sim -0.8$), the specific PFT parameters E_{max} and z_1 are fairly correlated ($r \sim 0.6$), while the non-specific parameters g_m and α_m are moderately correlated ($r \sim 0.5$). Moreover, there are also other pairs of weakly correlated parameters such as θ and b_{C3} ($r \sim 0.3$), the minimum canopy conductance of grass (g_{minG}) and g_{minT} ($r \sim 0.3$), g_{minG} and PT ($r \sim 0.3$), g_{minT} and a_m ($r \sim 0.3$). Finally, weak anti-correlations ($r \sim -0.3$) occur between the three pairs θ - PT , α_a - λ_m and α_{C3} - λ_m .

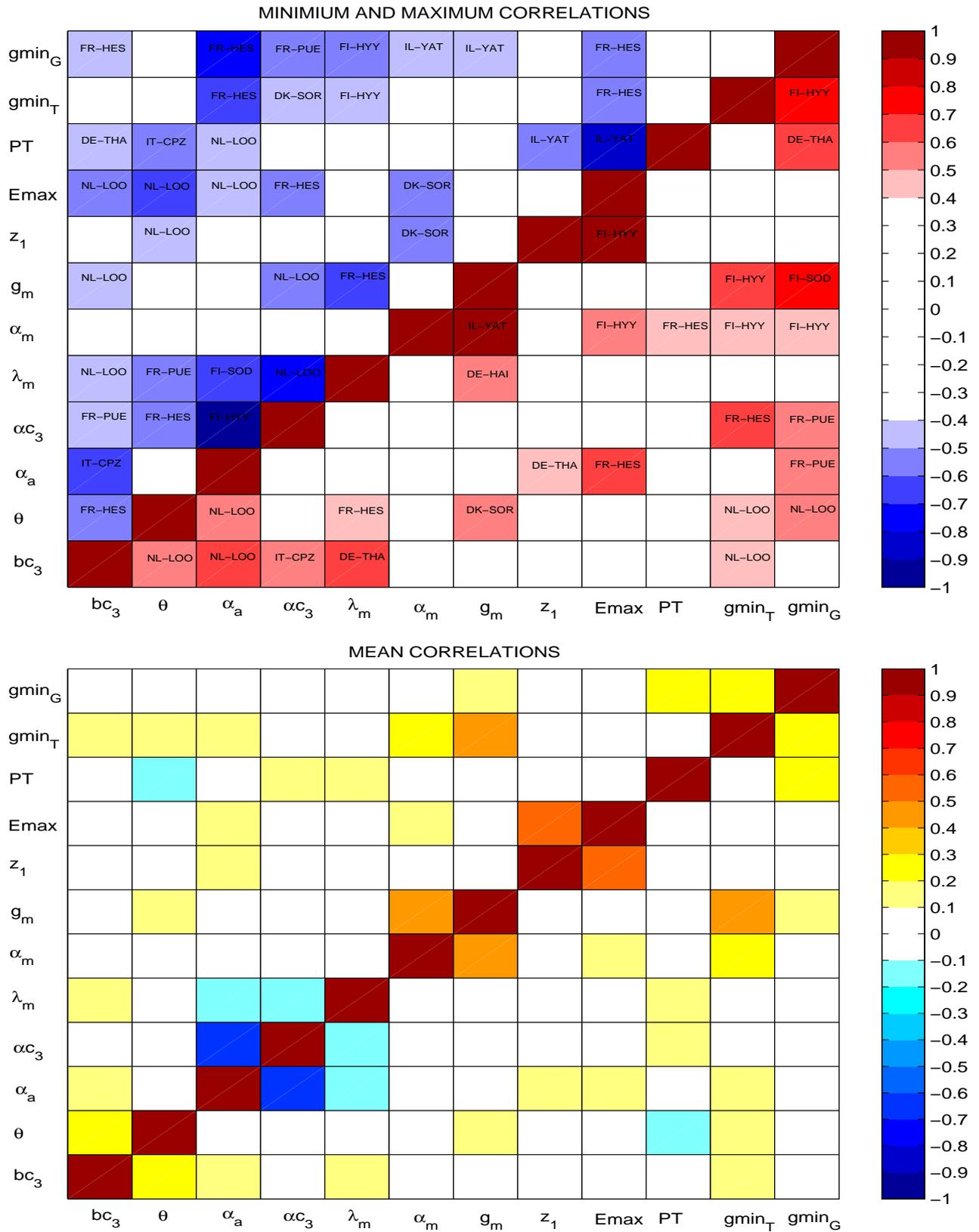


Figure 3.28. The first panel shows the maximum correlations and anti-correlations for any couple of parameters; such extreme values have been found searching over all the ten stations. In addition, each filled box points out the station where such maximum occurs. The second panel shows the mean correlation for any couple of parameters, the mean being computed by averaging over the ten stations.

3.3.6 Validation against eddy data for year 2003

In order to assess the performances of the model we used the final-optimized parameters to simulate water and carbon fluxes for the year 2003. For each station we computed R^2 and RMSE between data and model, using both prior and final-optimized parameters to perform the simulations.

Both for R^2 and RMSE the relative increments have been computed according to Equations (3.20) and (3.21) respectively, with the difference that the sum refers only to the year 2003. The results are shown in *Figure 3.29*. The positive relative differences indicate that for GPP, ET and TER there is an improvement in the model when we employ the final-optimized parameters, while for NEE there is a reduction of performances in simulating this flux; these reductions also occurred in *Figure 3.14* but only for R^2 . However, if we compare data in *Figure 3.29* with the same of *Figure 3.14* we see that for 2003 we got lower values both for R^2 and RMSE increments: this result confirms how data assimilation is a useful method to improve the model performances.

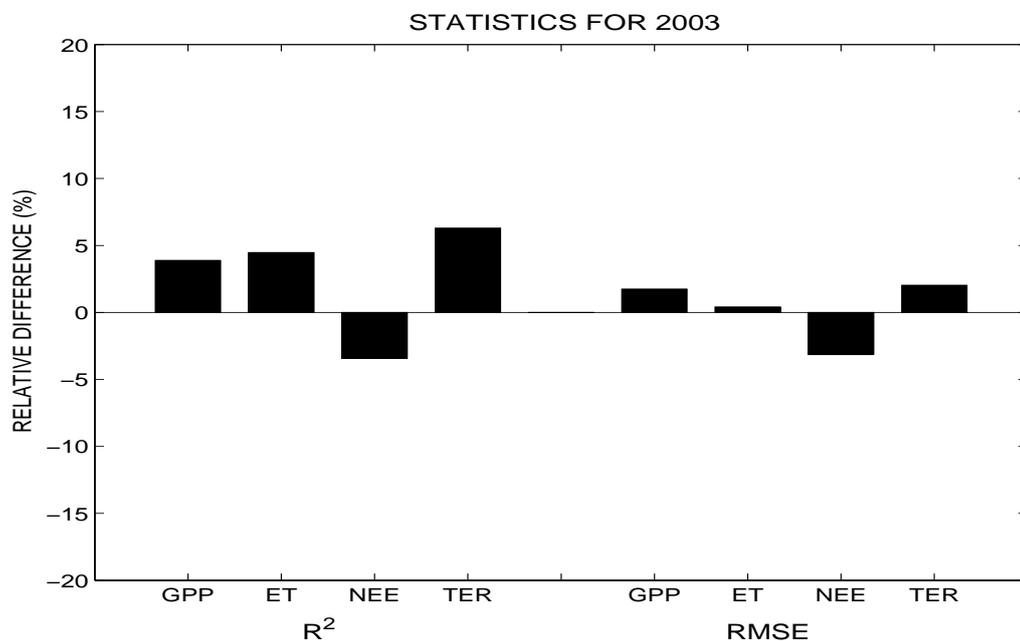


Figure 3.29. Relative increments between prior and posterior (final-optimized) squared correlation coefficients (R^2) and root mean squared error (RMSE).

The above model validation, however, have been performed just on those sites involved in the data assimilation; thus, to evaluate the model performances, some test sites not used during the

model inversion have been chosen to compare GPP and ET fluxes computed using both prior and posterior parameterization. Results are shown in *Figure 3.30*. Generally, the optimized version of LPJ reproduces water and carbon fluxes slightly better than the prior version; in both cases, however, LPJ simulates correctly the phase of the seasonal cycle and also the amplitude of these fluxes.

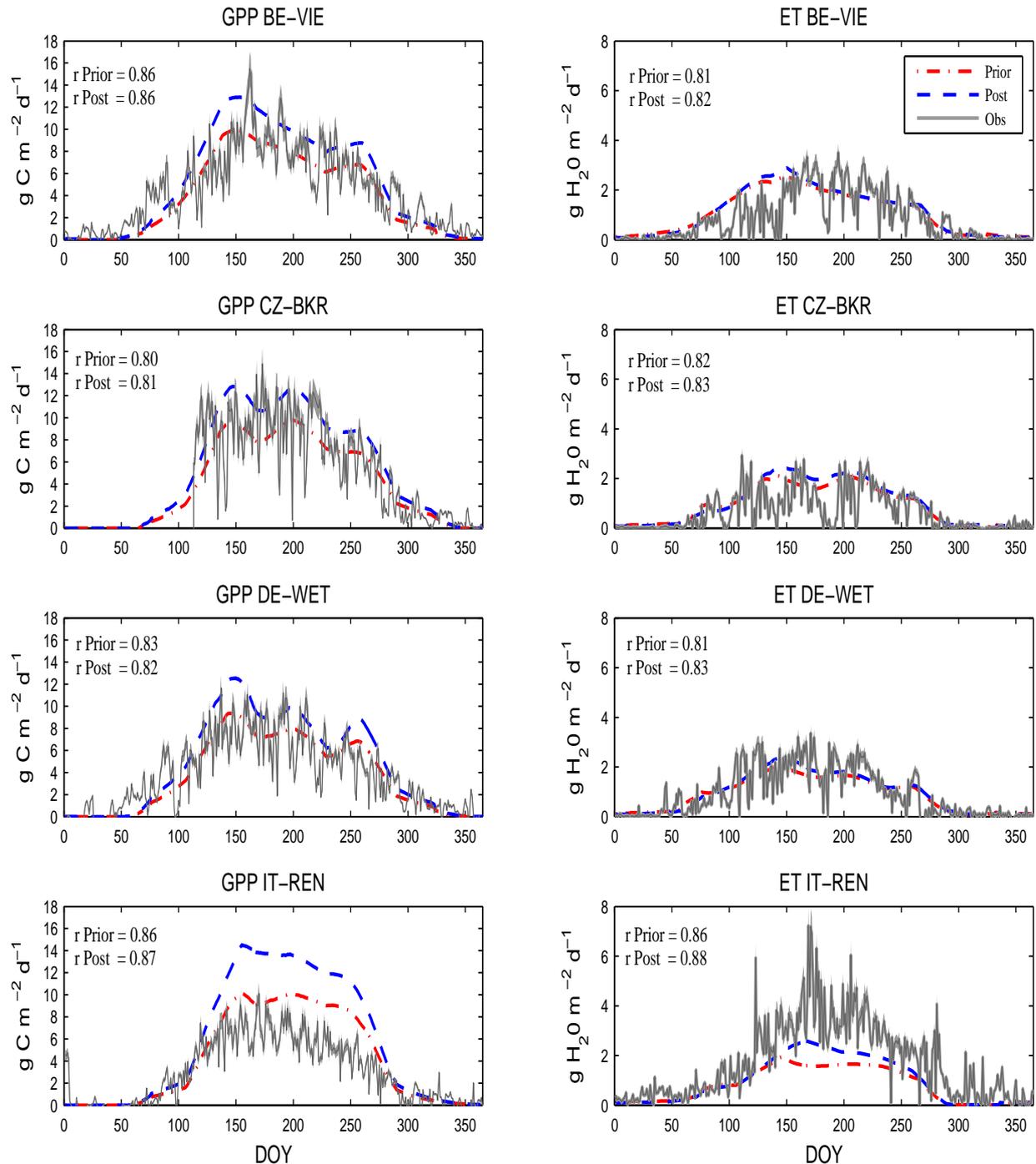


Figure 3.30. Comparison of observed and simulated water and carbon fluxes in four CarboEurope sites:

Finally, for year 2003 we carried out a last simulation in order to evaluate the performances of the model with respect to the summer heat wave. In *Figure 3.31*, is mapped the European-wide productivity anomaly as the difference between the control run (prior) and posterior (final-optimized) by averaging over the summer season June, July and August (JJA).

Using the final-optimized parameter values, the model captures the productivity changes in the Mediterranean basin related to the heat wave and the drought effect. *Figure 3.31* also shows how GPP increases in Northern Europe in response to moderate warming and no marked water deficits. These changes in GPP agree with the eddy data and satellite observations and are of the same magnitude as described in other papers [Ciais *et al.*, 2005; Reichstein *et al.*, 2005].

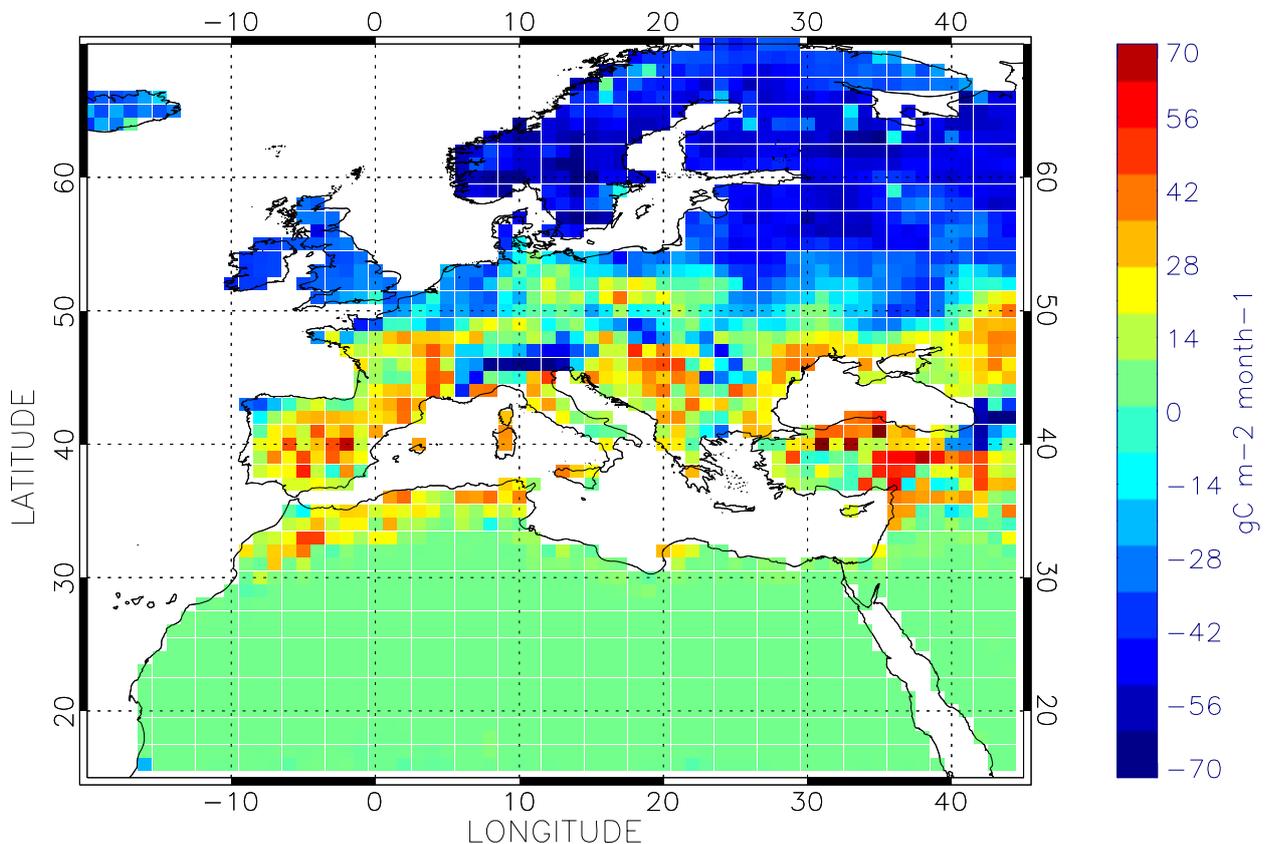


Figure 3.31. European-wide anomaly of GPP during the summer 2003. The figure represents the difference, averaged over JJA, between prior and posterior, for the summer 2003.

From the modeling point of view the 2003 heat wave can be regarded as a proxy of future climate that will be warmer and with more extremes [Schär *et al.*, 2004]. Hence, a successful modeling of the short-term effects on the ecosystems carbon and water on the terrestrial ecosystem is a necessary (but not sufficient) condition for confidence of future model predictions [Reichstein *et al.*, 2007].

3.3.7 Global effects of the new parameterization

The aim of this section is to understand how the new parameterization affects water and carbon fluxes on the whole globe. For such purpose, we used both prior and final-optimized parameters to initialize globally the model. We performed three different simulations: in the first one (henceforth “*Prior*”) we used the prior parameter values to spin up the model and to compute fluxes till 2003, in the second case (henceforth “*No-Spin up*”) we used prior parameter for the spin up and to run the model till 1993, and next final-optimized parameters to perform the simulation from 1993 until 2003, while in the latter case (henceforth “*Spin up*”) we used final-optimized parameters for both spin up and simulation. During the spin up phase of the model, the initialization of carbon soil pools has not reached unstable values, so we verified that the new parameterization does not force all the model variables to become unrealistic. To understand how fluxes differ between prior and posterior case, we should first of all analyze the prior case and next we could look at the anomaly of posterior fluxes with respect to the prior case. Since LPJ-DGVM has been comprehensively validated for terrestrial carbon and water exchanges and vegetation distribution here we briefly give only few comments about prior results. The geographical distribution of net primary production, net ecosystem exchange, evapotranspiration and soil carbon stocks (for the prior model simulation) are displayed in *Figure 3.32* as averages of the time-period 1993-2003. For the present climate (1993-2003), LPJ simulates global totals vegetation, soil and litter carbon are 845 ± 13 , 1400 ± 1.3 and 248 ± 0.8 PgC, respectively (*Table 3.11*). All of these estimates are within the range of other model studies [*Foley et al., 1996; Kucharik et al., 2000; Sitch et al., 2003; Krinner et al., 2005; Zeng et al., 2005*], although vegetation carbon is at the high end of the range. According to *Sitch et al.* [2003], the NPP is estimated by the model to be 64 PgC for the period 1901-1970, increasing to an average of 70 ± 1.8 for the present climate. Also the simulated global NPP is on the high end of the range of 44.4–66.3 PgC yr⁻¹ computed by the terrestrial biogeochemistry models participating in the Potsdam NPP Intercomparison Project [*Cramer et al., 1999*]. A reason for part of this difference may be that 1980s climatology and CO₂ levels were used here, while most of the simulations carried out in the NPP model intercomparison exercise used the 1930–1961 climate input, which is slightly cooler in the Northern Hemisphere [*Sitch et al., 2003; Krinner et al., 2005*], and a slightly lower atmospheric CO₂ concentration (340 ppmv). It might also be in part due to the fact that nitrogen limitation, a key control on productivity in many ecosystems, particularly in the high latitudes, is not taken into account in LPJ (however, most of the models participating in the intercomparison project do not model the nitrogen cycle either) [*Cramer et al., 1999*].

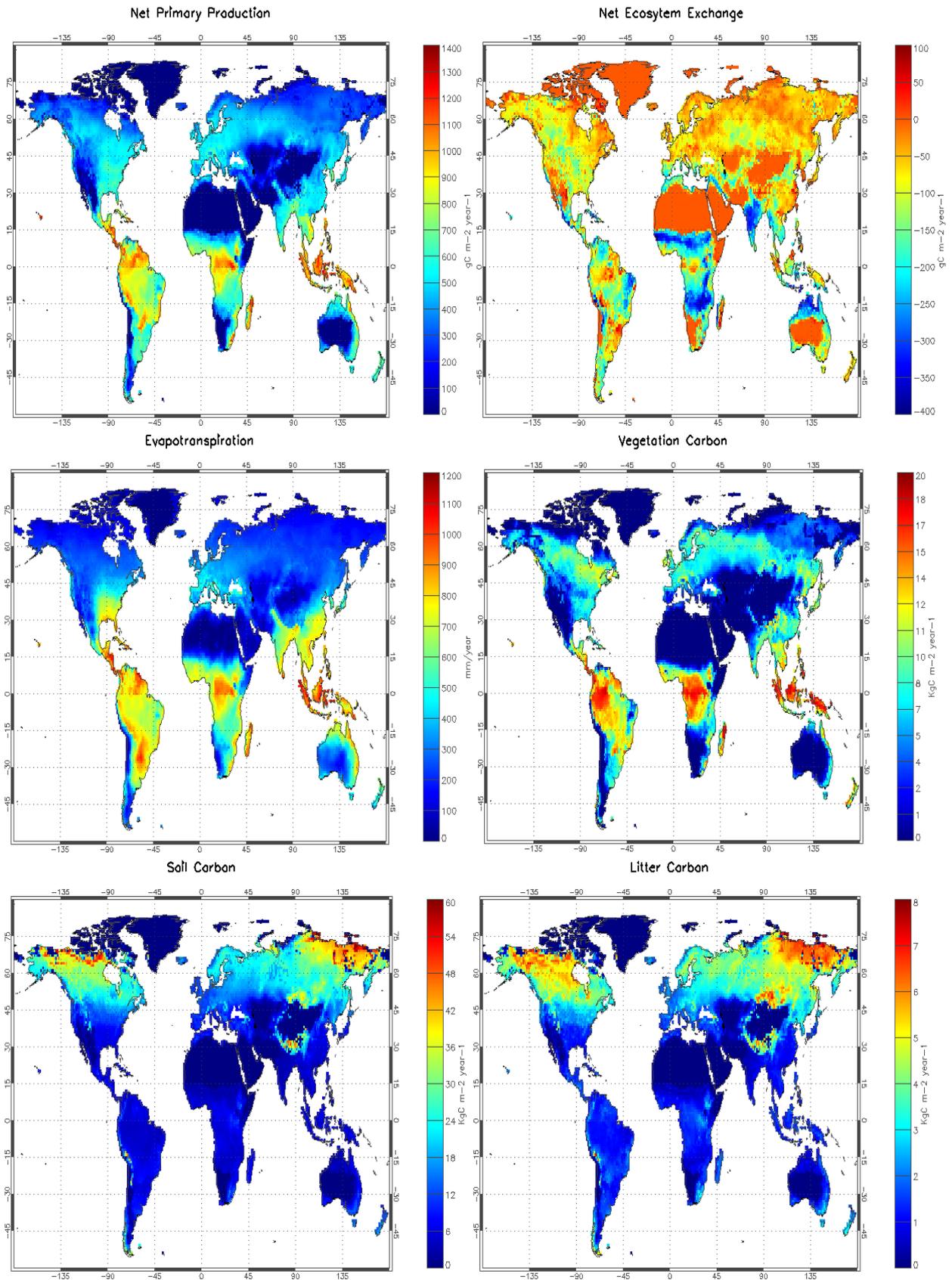


Figure 3.32. Geographical distribution of model simulated (Prior) water and carbon fluxes and soil pools averaged for the time period 1993-2003.

All these results are already published in *Sitch et al.* [2003] for such reason they will not be discussed any further in the following. Henceforth we will compare these prior results with the same result we got using the new parameterization for the whole globe. The anomalies are shown in *Figure 3.33* as differences between posterior (*No-Spin up*) and prior fluxes.

Looking at the six panels of *Figure 3.33* we infer that the new parameterization yields a remarkable increase of the amount carbon assimilated by the vegetation and a slight decrease of the evapotranspiration at midlatitudes. The enhanced carbon stored by mean the NPP increases, in turn, both the amount of living biomass (vegetation carbon) and the amount of death biomass that reaches the soil and goes to build up the litter and the soil carbon pools (*Figure 3.33*); globally, the *No-Spin up* simulation generates a stronger sink of carbon than *Prior* one (*Figure 3.33*).

The NPP increases roughly of 12 PgC in the period 1993-2003 for both the *No-Spin up* and the *Spin up* simulations (*Table 3.11*), and also the uncertainty associated to the NPP is increased (*Table 3.11*). Since *Prior* NPP is on the high end of the range computed in the intercomparison project [*Cramer et al.*, 1999] this result highlights how the new parameterization, computed for the European forests, seems to be not applicable to the whole world. It is also noteworthy that the two simulations that make use of the new parameterization simulate much more vegetation and litter carbon than *Prior* case particularly at mid-high latitudes. Results integrated on all the model grid cells (67420 points) are shown in *Table 3.11*; it quantitatively points out the changes in carbon pools between the three simulations.

Simulation (1993-2003)	NPP (PgC)	VEGETATION CARBON (PgC)	SOIL CARBON (PgC)	LITTER CARBON (PgC)
<i>Prior</i>	70±1.8	845±13	1400±1.3	248±0.8
<i>No-Spin up</i>	82±2.2	959±16	1500±4.3	306±1
<i>Spin up</i>	82±2.2	965±15	1694±2.1	309±0.9

Table 3.11. Cumulative changes in terrestrial vegetation carbon and soil pools (averaged on the time period 1993-2003) under different parameterizations and model initialization.

The *No-Spin up* simulation compute 114, 100 and 58 PgC more than *Prior* simulation for vegetation, soil and litter carbon, respectively, while if we consider the *Spin up* simulation the bias is increased too. It is also noteworthy that also the standard deviation associated to the last two simulations is enhanced for all the variables taken into account (*Table 3.11*).

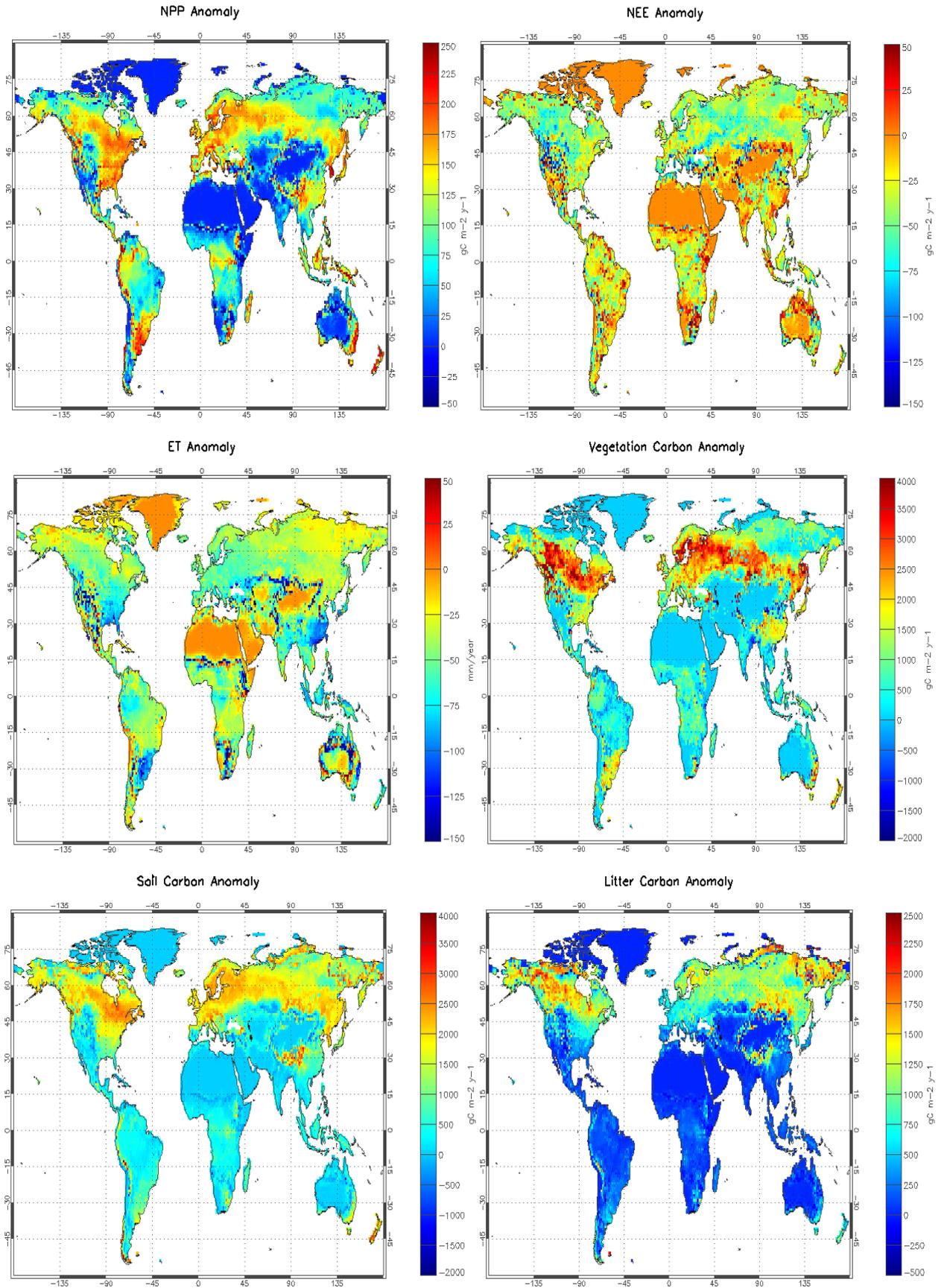


Figure 3.33. Geographical distribution of anomalies (No Spin up-Prior) for water and carbon fluxes and soil pools averaged for the time period 1993-2003.

In the *No-Spin up* and *Spin up* simulations the increase of NPP and, hence, of carbon pools at mid-high latitudes seems to be directly related to the dominant vegetation present at the given grid cells. *Figure 3.34* shows the potential natural vegetation simulated by the model for the present climate; all the three simulations gave the same distribution of the dominant vegetation. The dominant vegetation in each grid cell is represented by the PFT with the highest simulated area cover (i.e. fractional percentage coverage).

Here we can discern the large area of boreal evergreen forest in Canada and Northern Eurasia, the boreal deciduous forests in Siberia, and the transition into temperate ecosystems of North America, western Europe and China. Moving further south LPJ simulates the transition from drought deciduous forests of the sub-tropics, having distinct wet and dry seasons, into the evergreen rainforests around the equator. LPJ is moderately successful in simulating non-wooded areas, including grasslands. LPJ simulates northern tundra, and grasslands in dry areas including western USA, southern Europe and central Asia. The dominant PFT in the southern Russian steppe is modelled incorrectly as temperate summergreen tree.

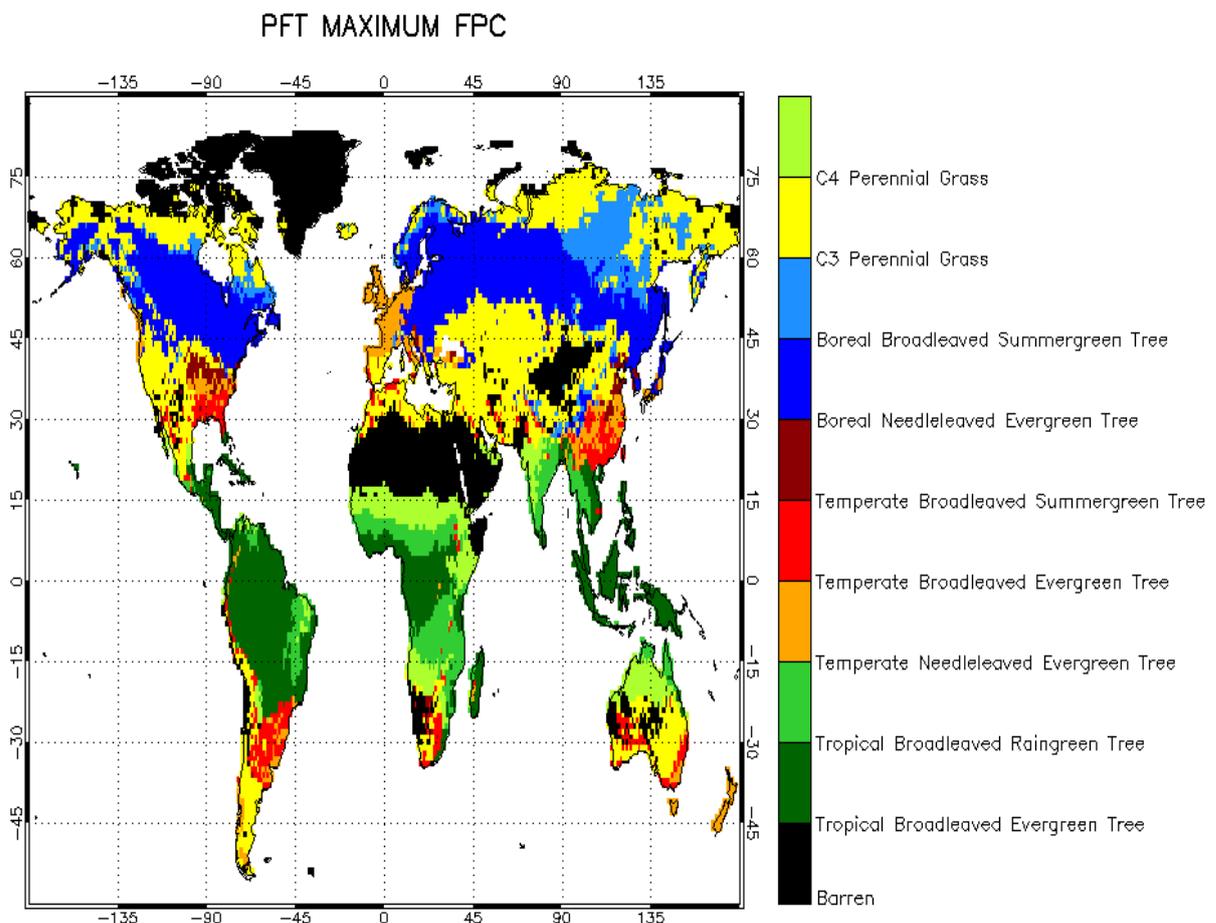


Figure 3.34. Simulated dominant PFT (PFT with maximum percentage coverage for each grid cell).

Looking at *Figure 3.33* and *Figure 3.34* it is noteworthy that the areas where the maximum changes of carbon pools occur are those where the BNE is the dominant vegetation; this might point out that the posterior values valid for the PFT-specific parameters controlling the boreal needle-leaved evergreen trees could yield an overestimation of the carbon pools. This result, however is still under examination, and in the future we expect to conduct further studies to assess the relative variation of water and carbon fluxes for each PFT due to different parameterizations. Finally to assess how the fluxes vary locally between the three simulation, we divided the whole globe in six sub-regions, as described in *Figure 3.35*.

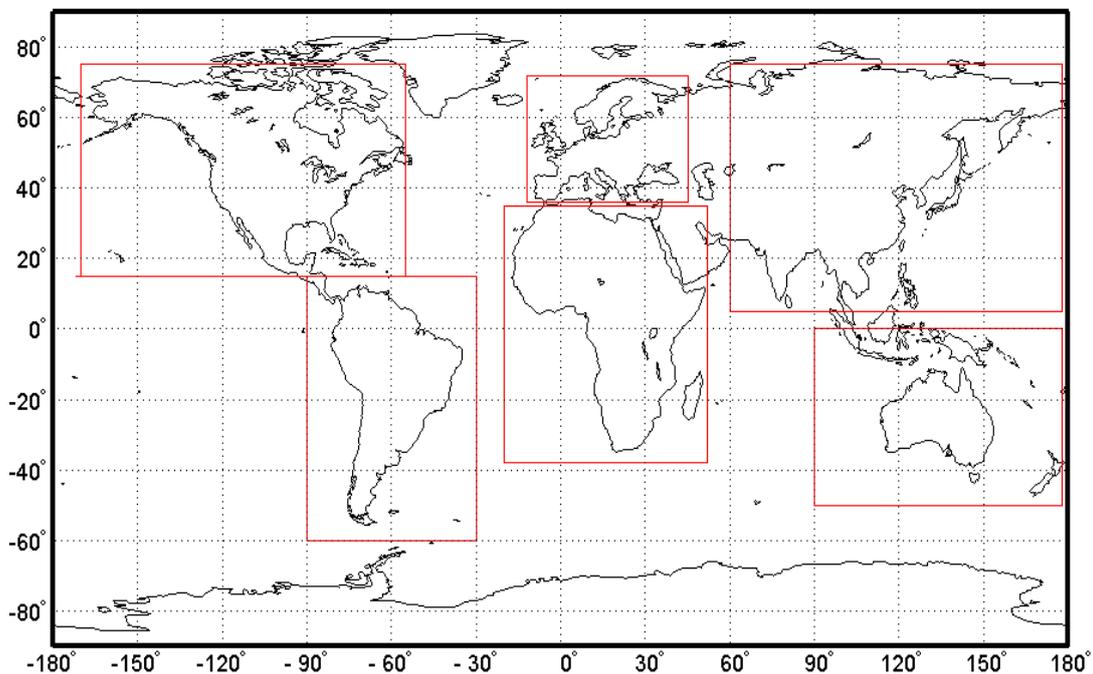


Figure 3.35. Six world sub-areas and their bounds (latitude-longitude): South America (-60, 15; -90, -30), North America (15, 75; -90, -30), Africa (-38, 35; -20, 52), Europe (36, 72; -12, 45), Asia (5, 75; 60, 178) and Oceania (0, -50; 90, 178).

In each sub-area we computed water and carbon fluxes; the time-series of the three simulations are plotted in the figures below.

As for NPP (*Figure 3.36*), no significant differences have been found in the six sub-domains between *No-Spin up* and *Spin up* simulations; besides, these last two runs yield higher fluxes with respect to the *Prior* simulation. The main differences between the prior and the posterior results are in Europe and North America, while in Africa and Oceania only weak variations occur. This result confirm how the new specific parameters valid for BNE plant functional type yield a remarkable dissimilarity in term of plant assimilation ($\sim 120 \text{ gC m}^{-2} \text{ y}^{-1}$ in Europe and $\sim 100 \text{ gC m}^{-2} \text{ y}^{-1}$ in North America).

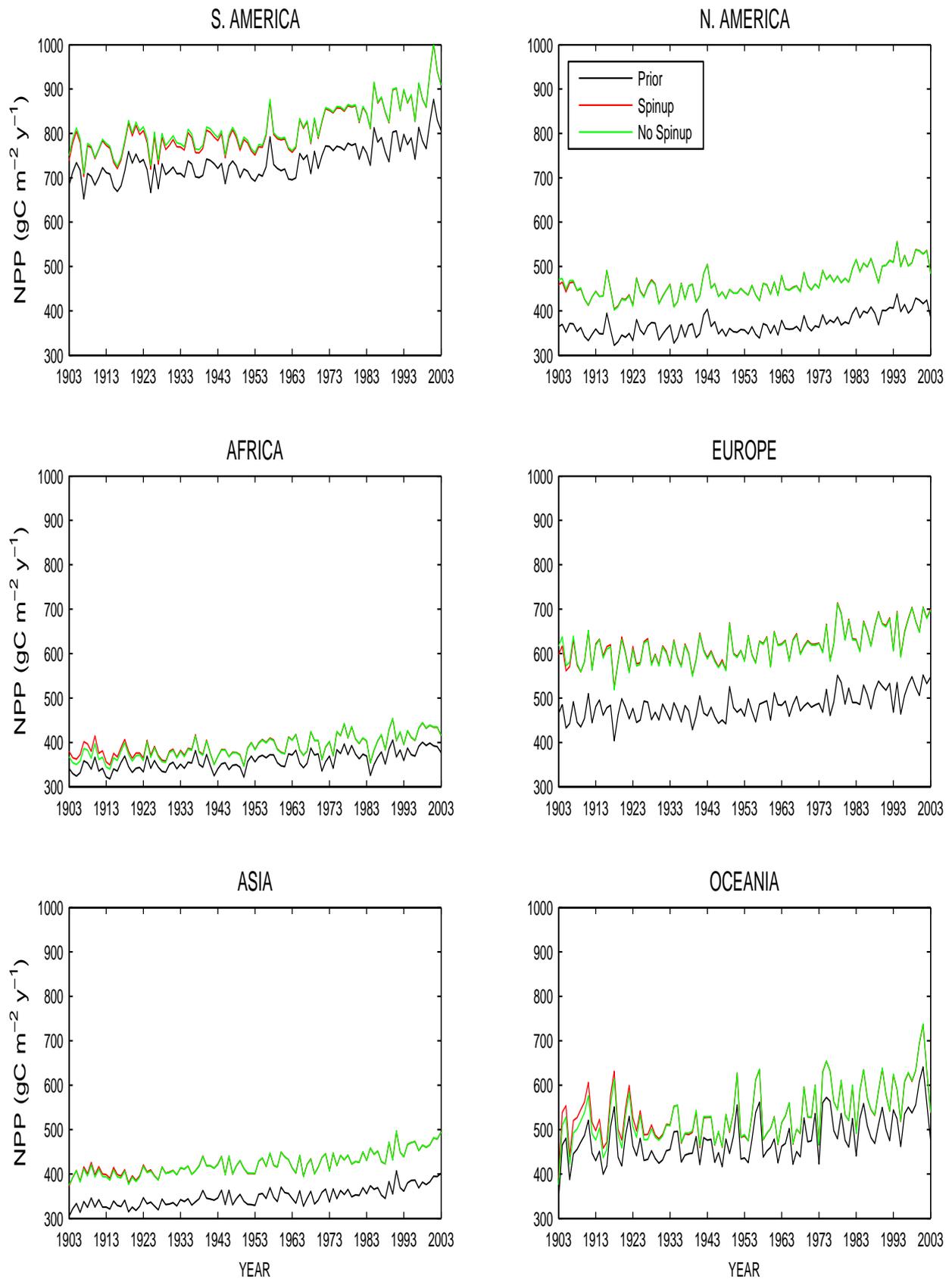


Figure 3.36. Interannual variability of NPP from various regions of the world and for the three different experiments (Prior, No-Spin up and Spin up).

Taking into account the evapotranspiration, the only relevant differences from the prior and the posterior simulations occur in South America and Europe, while, unlike the NPP, in North America no significant difference between the three simulations have been found. As already shown in *Table 3.11*, the new parameterization yields broad variation in the carbon pools, and in particular the carbon soil shows the larger variation in all the six sub-regions (figure not shown).

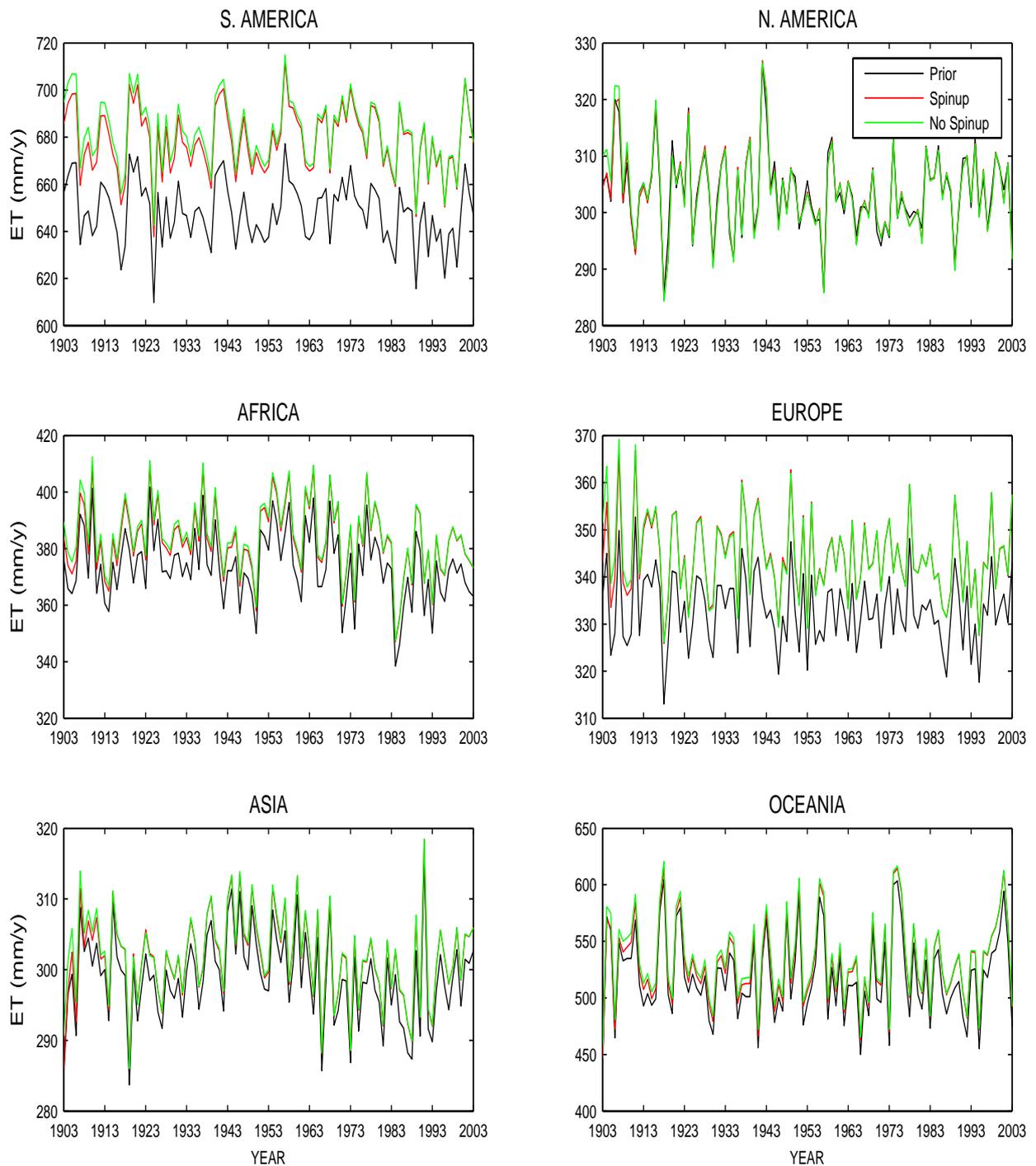


Figure 3.37. As Figure 3.36 but now is considered the ET variable.

3.4 DISCUSSIONS AND CONCLUSIONS

Parameter optimization of terrestrial ecosystem models provides a direct method for assessing model parameter values in an optimal way so that models produce results that are consistent with the observed data at site level. The parameters of ecological models have generally been derived from small numbers of measurements at the plant scale or are estimated in laboratories. However, parameters determined at one scale lead to incorrect predictions at other scales if there is a nonlinear relationship between model parameters and predicted fluxes [Wang *et al.*, 2007]. In addition, also the climatic conditions, under which parameters values have been computed are not constant, but may change in time (e.g. atmospheric CO₂ concentration). To overcome this problem inverse algorithms have been developed to assimilate observations at a range of temporal and spatial scales. In this context, the ultimate aim of this study was to relate the performances of LPJ to existing observations in Europe and retrieve the PDF of model parameters controlling water and carbon fluxes. This was achieved by a synthesis of observational data with model results, applying data-assimilation methods.

This study has demonstrated how data assimilation provides a powerful tool for analyzing ecosystem processes and it might help to improve our understanding about carbon and water exchange between land ecosystem and atmosphere. Data alone, in fact, are often insufficient for this task or, at least, problematical because of gaps in time series, or methodological uncertainties [Williams *et al.*, 2005]. Gap-filling techniques are generally highly statistical and may produce inaccurate results. A conventional modeling approach to ecosystem analysis may also be problematical because TEMs and dynamic global vegetation models are often tuned to deliver estimates on wide regions, and that means that the models provide little extra information at local scale. In fact, these models have large numbers of parameters valid for several different combinations of biomes, so these parameters, as demonstrated above, could be tuned to match the measurements at stand site. Data assimilation, as we have demonstrated here, is useful to solve many of these problems. In the last decade, data-assimilation techniques were widely utilized to identify the optimal parameter values of different terrestrial ecosystem models [Braswell *et al.*, 2005; Knorr and Kattge, 2005; Santaren *et al.*, 2007; Wang *et al.*, 2007]. Optimal sets of parameter values were retrieved by minimizing a cost function which described least square misfits between observations and model results. Measurements from time-series studies, mainly collected in the CarboEurope or Fluxnet programs, entered in the data-assimilative investigations.

In this thesis we estimated twelve model parameters in the LPJ dynamic vegetation model constrained by eddy covariance data of carbon and water fluxes measured in ten different

CarboEurope sites. The sites were chosen in order to represent the main European forests represented into the model in terms of four different plant functional types. The measurements used during the inversion covered a period of 2000 through 2002, while 2003 was used to test model performances with respect to the new parameterization.

Two different optimisation methods were developed to assimilate eddy observations into LPJ. One method relied on an adjoint model which provided gradient information of the cost function for parameter optimisation with a gradient search algorithm. The second method was based on the usage of a stochastic search algorithm, the Metropolis algorithm, based on the Monte Carlo Markov Chain method.

Unlike gradient methods, we found that in LPJ the Metropolis algorithm works very well in the multiple constraint approach thanks to the ability to deal with non-linear problems. In other words, during the optimizations the gradient methods was not able to converge to a local/global minimum due to the complex shape of the cost function, while with MCMC we reached the convergence in all the simulations we performed. In addition the MCMC method allowed sampling the complete posterior PDF even if it is generally highly CPU-expensive and required long simulations. For such reason, we performed several Bayesian inversions of LPJ model using the Metropolis algorithm.

We performed two kind of optimization differing on the way observation are used: in one case we ran the assimilation using all the data in one step from 2000 through 2002; in such case the inversion make time-constant parameters. In the other case, we used three single year of observation in order to produce parameters that are able to vary in time.

While previous published studies have shown the potential of model inversion against eddy covariance data and have emphasized the importance of data error characterization or model performances, in this thesis we looked also at spatial and temporal variability of parameters controlling carbon and water fluxes.

The optimization procedure we performed helps to retrieve a better knowledge about some LPJ ecophysiological parameters and uncertainties of the four different PFTs we take into account. In addition, the calculation of the error covariance matrix of the optimal parameters provides a way of evaluating the sensitivity of the model output to the model parameters and assessing the correlations between the various parameters in the model.

By means of a qualitative description, we found a relevant time-dependence of some parameters: by assimilating fluxes for each year separately we showed how there is a temporal variability in the results. In other inversion studies for some deciduous forests, *Wang et al.* [2007] found some parameters to vary seasonally as leaves develop. To our knowledge, seasonal and interannual variation of parameters have not been taken into account in any land-surface schemes used in

global climate modeling so far. This may result in systematic bias in the modelled net carbon fluxes from deciduous forest and predicted seasonal variation of atmospheric CO₂ concentration over the land surface in mid-latitude regions [Wang *et al.*, 2007].

On the other hand, we expected that non-specific PFT parameters of the Farquhar model of photosynthesis, and of water uptake and potential evapotranspiration change only slightly between different PFTs and within sites with the same PFT because dynamic global vegetation models should be able to simulate correctly fluxes for all the biomes of the world using the same non specific PFT parameter values. Nevertheless we point out that some parameters have a low or modest interstation variability, while others show a larger interstation variability. However, ANOVA test showed how the spatial variability is not significant and the parameters of the Farquhar model are PFT independent.

The high spatial parameters variability highlight how the model and parameterization is not designed to deliver good estimates at single stand sites but on larger regions. In the development of global land-surface models, one of the major difficulties is to ensure that the models are applicable to a wide range of terrestrial ecosystems, ranging from desert to arctic biomes, under the present and future climate conditions [Wang *et al.*, 2001]. Obviously, at some sites the local influence due to particular soil or climate conditions may be very high and the model can only cope with it by changing posterior parameters. On larger regions or for the whole globe, however, the averages over several sites are capable producing reasonable estimates.

We also showed how the final-optimized values are very close for almost all parameters to the prior values for either assimilation schemes. For the parameters where a large interstation variability have been observed, this depends on the fact that differences occur randomly in both ways and generally tend to compensate with each other, while where a low variability occurs we expected a posterior value almost equal to the prior one.

The *t*-test and the relative increment of correlation coefficient and RMSE showed how data assimilation is a useful tool to improve the ability of the model to simulates correctly water and carbon fluxes at local scale: the inversion successfully matches the observed seasonal cycle of the diverse fluxes, and corrects for the prior misfit to day-time GPP and ET. In addition, extreme events such as 2003 summer heat wave, were correctly captured.

Anyway, optimization could force other model variables to become unrealistic; here we showed how assimilation also decrease the model-data misfit for NEE and TER even if these have not been taken into account during the optimization.

Performance of data assimilation tended to be poorer at Mediterranean sites than in the other central on Northern Europe sites, with representations of the mechanisms of response of plant physiology and allocation, and of microbial activity, to soil water deficits being identified as

possible causes of model-data mismatches. This implies that there may be some structural deficiencies in the model, and data assimilation partially overcome such problem slightly improving model-data fit.

In any case other inversion studies are needed in order to identify the right values for such parameters and their corresponding uncertainty. It is convenient to collect more data and increase the number of station where to run optimizations and possibly apply both Metropolis and other inversion methods by running the optimization with a reduced set of parameters, since we demonstrated, in this study, that most of them have posterior values close to their prior ones. It could be also useful to use different prior uncertainties values to check result coherence.

IV

EFFECTS OF LAND COVER CHANGES ON CLIMATE OVER EURO MEDITERRANEAN AREA

4.1 INTRODUCTION

Land use and Land Cover Changes (LCC) affect the local, regional and global climate system through biogeophysical and biogeochemical processes [*Pielke et al., 2007*] that modify both surface-atmosphere exchanges of momentum, energy and greenhouse gases and surface roughness [*Xue and Shukla, 1993; Xue and Shukla, 1996; Fahey and Jackson, 1997; Betts, 2001; Pitman, 2003; Nosetto et al., 2005; Pielke et al., 2007*]

The feedback mechanisms between the land surface and the atmosphere have been increasingly investigated during latter decade due to the increasing computational power. Therefore, in order to study the potential impacts of LCC on local climate, the simulations performed by general circulation models have been complemented by the use of regional climate models: in fact, the coarse resolution of the GCMs limits their capability to capture mesoscale features that play a pivotal role in regional dynamics [*Giorgi and Mearns, 1991; Gaertner et al., 2001; Seneviratne et al., 2006*].

The vegetation cover, in particularly over Europe, has deeply changed in the last centuries due to the human activities that converted forests into arable land or pasture and it has influenced considerably regional climate [*Reale and Shukla, 2000*].

Potential resulting impacts from vegetation changes include changes in the radiation budget via the surface albedo and changes in the hydrological cycle in terms of evaporation, precipitation

and runoff. Considering model sensitivity studies, it has been observed that the strength and regional variation of such impacts, however, depend on the atmospheric model that is used [Dumenil-Gates and Ließ, 2001].

In continental midlatitudes soil plays a similar role to that of the ocean, but instead to store heat, it store water during the winter and during the summer it moisten the atmosphere by evapotranspiration [Heck et al., 2001; Koster et al., 2004]. Some GCM studies suggest that in Europe the summer soil moisture anomalies do not have a substantial impact on precipitation [Koster et al., 2004; Seneviratne et al., 2006]. Nevertheless, regional simulations emphasize the importance of soil moisture-temperature and soil moisture-precipitation feedbacks in influencing summer climate variability in Europe [Seneviratne et al., 2006]. These results highlight the importance to use both global and regional models; the later could take into account mesoscale processes which impact on the feedbacks at local-regional scale.

Several attempts to study and quantify the effect of LCC on climate exist for Europe. Dumenil-Gates and Ließ [2001] found a slight cooling at the surface (1 °C) and reduced precipitation during the summer as a result of less evapotranspiration of plants in a deforestation experiment, while they found opposite results for the afforestation experiment. As for winter season, no significant signal have been found due to the strong influence of the mid-latitude baroclinic disturbances [Dumenil-Gates and Ließ, 2001]. Sanchez et al. [2007] performed a simulation where they substituted trees with grass: in such case, they found a significant decrease of summer precipitation up to 3 mm/day due to the less evapotranspiration in the grass simulation and an increased surface temperature up to 3 °C. In Heck et al. [2001] the afforestation cause a maximum cooling of 2 K during May, while during August a heating of 1 K have been observed; as for precipitation, no significant differences have been found. All these sensitivity experiments reveal different responses to LCC in terms of signal and amplitude.

The aim of this study is to corroborate the previous results and to explore new LCC patterns. A further objective of this study is to analyze the impact of LCC on weather extremes. The European heat wave of summer 2003 has received considerable attention both because of its potential link to larger-scale warming patterns (e.g., “global warming”) and the large non-harvest loss of life associated with it [Fouillet et al., 2006; Zaitchik et al., 2006].

Several studies found that this regional heat wave was quite unique in comparison to the instrumental climate record [Schar et al., 2004]. It is also been shown that the amplitude of the heat wave could be amplified by land surface preconditioning [Vautard et al., 2007; Ferranti and Viterbo, 2006]. For this purpose we explore the changes in regional climate as a consequence of both hypothetical anthropogenic deforestation, occurring mainly in the East Europe where there is a progressive substitution of natural forests with crops, and theoretical

spontaneous afforestation in the Euro-Mediterranean area caused by the potential abandonment of crops. To achieve this objective we impose different land covers as boundary conditions to an hydrostatic regional climate model (RegCM3).

4.2 MODEL, DATA, AND EXPERIMENTAL SETUP

4.2.1 Model description

In order to assess the potential impacts of LCC on climate we performed three different simulations conducted with a regional climate model; the experiments differ solely in the definition of land-use category in terms of type of vegetation cover and they represent just theoretical LCC.

The regional climate model used in the present analysis is RegCM3, a 3-dimensional, σ -coordinate, primitive equation model initially developed by *Giorgi et al.* [1990, 1993a, b] and then modified as discussed by *Giorgi and Mearns* [1999] and *Pal et al.* [2000]. The vertical σ -coordinate is used to follow the terrain elevation; this means that the lower grid levels follow the terrain while the upper surface is flatter.

RegCM3 is extensively used, for such reason here we provide only a brief description of the main features, while a more detailed description can be found in *Pal et al.* [2007].

We ran the model from 1981 to 2000 using the first year as spin up; a 20 years simulation ensure an enough time period to see how change surface climate with respect to land cover changes and a robust interpretation of results.

The model domain is centred around the Italy at 41°N and 15°W and is projected on a Lambert conformal grid covering almost all Europe (except northern Scandinavia and Iceland) and North Africa; the domain covers 160x150 grid points in the longitudinal and latitudinal directions respectively with an horizontal resolution of 30 km. At this fine resolution, the main topographic features of the domain are captured; the model domain and topography are shown in *Figure 4.1*. We set vertical coordinates to 18 sigma levels with the top at 50 hPa.

RegCM requires initial conditions and time-dependent lateral boundary conditions for the wind components, temperature, surface pressure, and water vapour. Terrestrial variables and three dimensional isobaric meteorological data are horizontally interpolated from a latitude-longitude mesh to a high-resolution domain on the chosen projection.

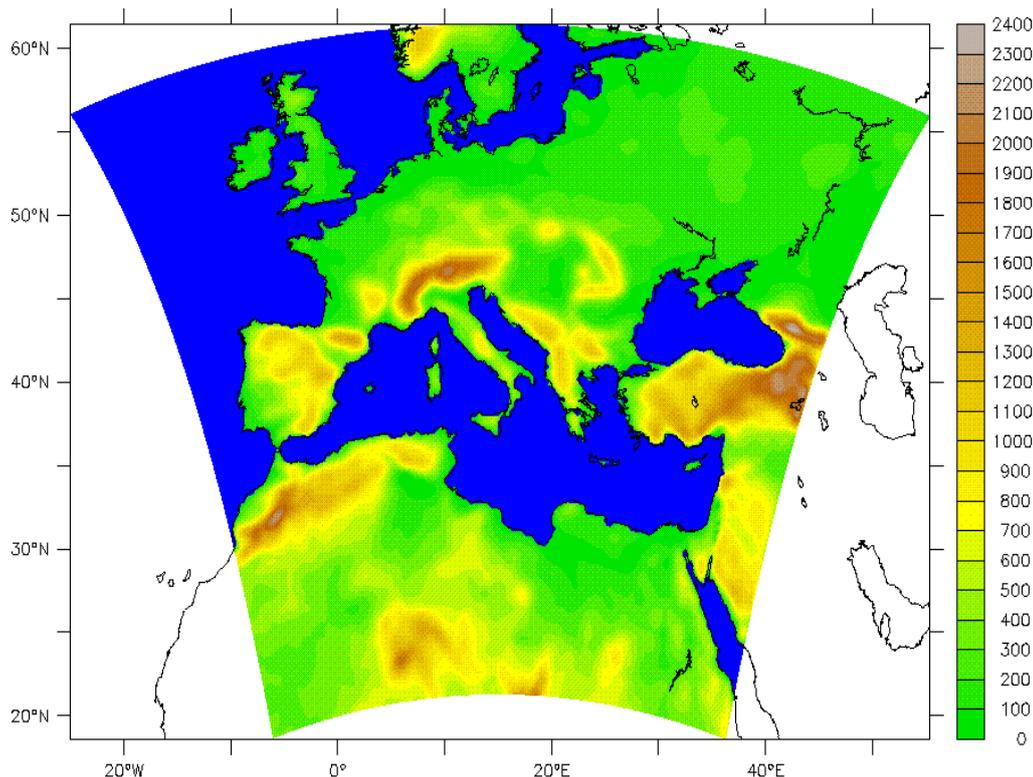


Figure 4.1. RegCM domain and topography (units are in meters) for the current experiments.

Lateral boundary conditions, required to run the model, are provided by the ECMWF reanalysis data [Uppala et al., 2005] with a resolution of $2.5^{\circ} \times 2.5^{\circ}$ that are interpolated at 6-hourly intervals via the relaxation method described by Giorgi et al. [1993b] into the model lateral buffer zone. RegCM3 also needs of Sea Surface Temperatures (SSTs) as boundary conditions; these are interpolated into model grid from the corresponding 1° NOAA fields [Reynolds et al., 2002]. This means that in the model the sea surface temperature is prescribed and the various physical phenomenon occurring into Planetary Boundary Layer (PBL) are unable to modify the SST. In nature, instead, changes in energy budget, as a consequence of changing in land cover, could cause a cooling or a warming of SSTs.

4.2.2 Potential vegetation

The exchange of heat, water and momentum between soil-vegetation and atmosphere is simulated in RegCM3 by the hydrological process model BATS (Biosphere–Atmosphere

Transfer Scheme, *Dickinson et al.*, [1993]). BATS divide the land surface into 18 types (*Figure 4.2*) and the soil in 12 types [*Dickinson et al.*, 1993]. These 18 classes of land cover are used to define a wide variety of land surface, hydrological and vegetation properties: each vegetation class, in fact, has associated a value of roughness length, albedo, LAI, rooting depth and the fraction of water extracted by the roots [*Dickinson et al.*, 1993]. As for 12 soil types, each of these has associated a value of porosity, minimal soil suction and the moisture content relative to saturation [*Dickinson et al.*, 1993].

The surface vegetation and landuse types are obtained from a global dataset derived by the United States Geological Survey (USGS) from satellite information [*Loveland et al.*, 1991]; at each grid element is assigned a dominant type of land cover [*Dickinson et al.*, 1993].

We performed three theoretical experiments to assess the impacts of LCC on regional climate. In the first experiment (CTL, *Figure 4.2a*), the USGS GLCC 10'×10' [*Loveland et al.*, 1991, 2000] land use dataset is used, which is representative of present vegetation cover over Europe. The second experiment (DEF, *Figure 4.2b*) consists in a deforestation scenario: starting from GLCC dataset, we substituted all the forests and the trees below 800 meters with crops/mixed farming. This experiment could be representative of the land cover changes occurring mainly over eastern Europe (*Figure 4.2b*) where there is a progressive deforestation leading to a substitution of forests with crops in order to produce biogas and biofuel. Deforestation in Europe also occur due to air pollution resulting from industries and transportation boundaries which is causing a lot of damage to natural species including trees, extended droughts and over grazing.

The third experiment (AFF, *Figure 4.2c*) considers a plausible evolution to a spontaneous afforestation situation caused by abandonment of crops and fields which lead to a natural recapture by forests of abandonment of arable land. Frequently farmers cease to use land because of high costs due to remoteness, difficult access, land of poor quality, steep slopes or high labour requirements, or where farmers' age and health prohibited use of land further from the farmstead [*Kobler et al.*, 2005].

Often a former mosaic of forests and farmland ended up wholly in forest. The absolute land use changes that we can detect may seem small, but in the long-term land use patterns may change drastically and these changes could modify local and regional climate.

As already stated above, at each grid element is assigned a dominant type of land cover; this mean that when we perform the sensitivity experiments changing the land cover we are assigning a new dominant vegetation at the given grid point; in other word, the forests (crops) in the DEF (AFF) experiment still remain in the given grid point but in fractional percentage cover less than the new dominant vegetation.

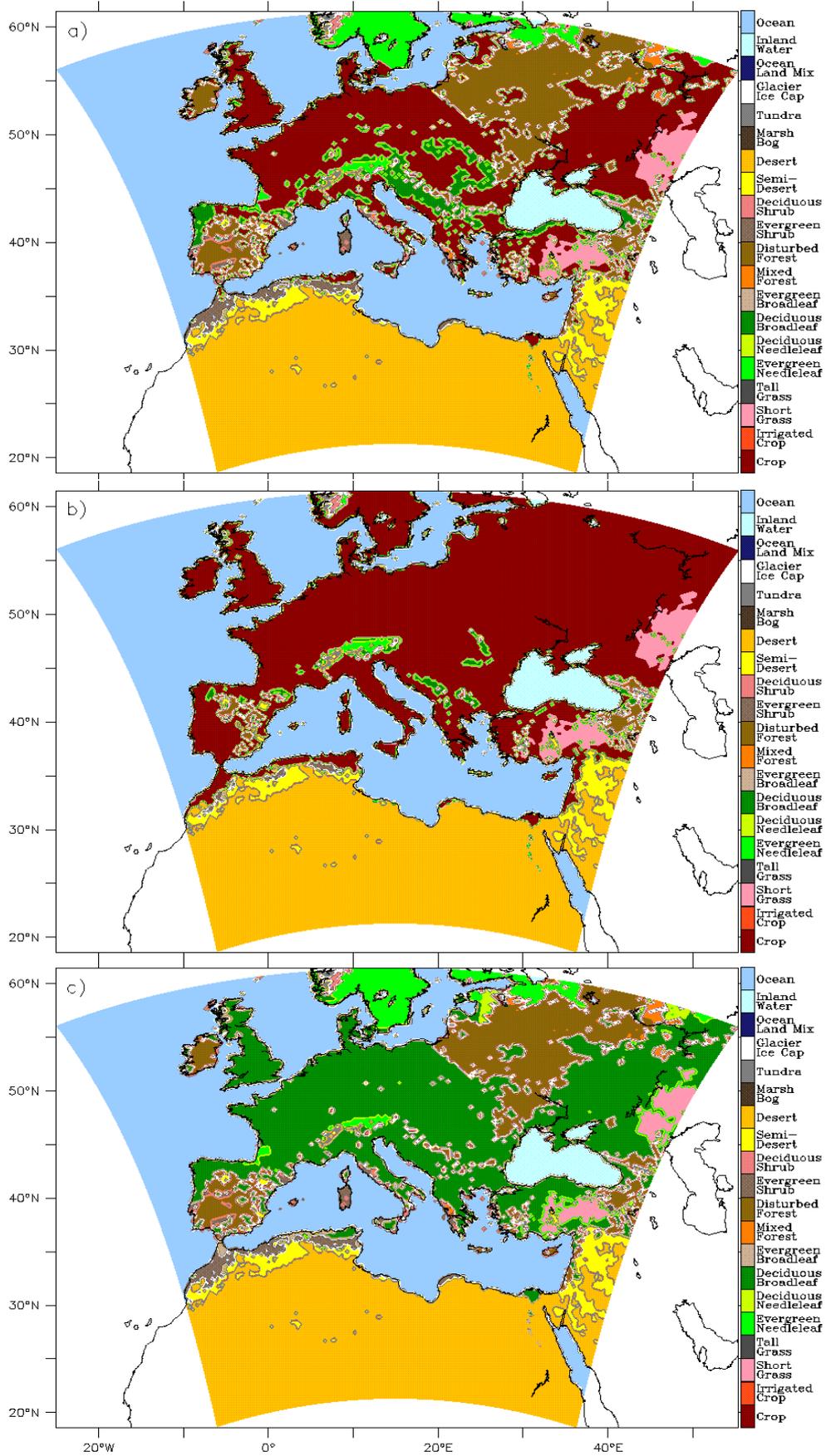


Figure 4.2. Land cover changes in Europe for control (a), deforestation (b) and afforestation (c) experiments.

4.3 RESULTS: IMPACTS OF LAND COVER CHANGE ON CLIMATE

4.3.1 Comparison of domain-averaged climate

We start by discussing the vegetation-induced changes in the evolution of the surface climate. At soil level, the faster response to LCC occurs modifying the surface energy budget. A quantitative comparison of time-domain averaged evapotranspiration (latent heat) and sensible heat between the CTL, DEF and AFF experiments is presented in *Table 4.1* for the winter (DJF) and summer season (JJA) respectively. For both variables taken into account all the values of DEF and AFF experiments are very close to the control run (*Table 4.1*). This result was partially expected because the mean over the whole domain and the entire period of simulation produce a smoothing for these variables and then a small difference from the CTL run.

SEASON	VARIABLE	CTL	DEF	AFF
DJF	EVAPOTRANSPIRATION (mm/day)	0.38±0.02	0.37±0.02	0.4±0.02
	SENSIBLE HEAT (W/m ²)	12.3±0.5	12.3±0.5	12.1±0.6
JJA	EVAPOTRANSPIRATION (mm/day)	1.7±0.09	1.7±0.09	1.6±0.09
	SENSIBLE HEAT (W/m ²)	80.6±2.5	78.2±2.3	82.6±2.1

Table 4.1. Comparison of domain-averaged heat fluxes (sensible and latent) between CTL, DEF and AFF experiments in the winter (DJF) and summer season (JJA), respectively. The time averages have been computed for the period 1982-2000; the first year of simulation (1981), in fact, have been considered as a spin up year, useful to initialize the soil moisture.

The temporal variation of domain-averaged means evapotranspiration and sensible heat for the three different experiments are shown in *Figure 4.3* for the winter and the summer season. In such case, the time series showed a similar behaviour between CTL, DEF and AFF simulations, with marked inter-annual variability.

The LCC modulates the amplitude of the sensible and latent heat fluxes leaving unchanged the inter-annual variability, which is mainly driven by the general circulation (i.e. lateral boundary conditions).

Furthermore, these results, in general, agree with other studies performed at mid latitude [Dumenil-Gates and Ließ, 2001; Heck et al., 2001; Gao et al, 2004; Sanchez et al., 2007].

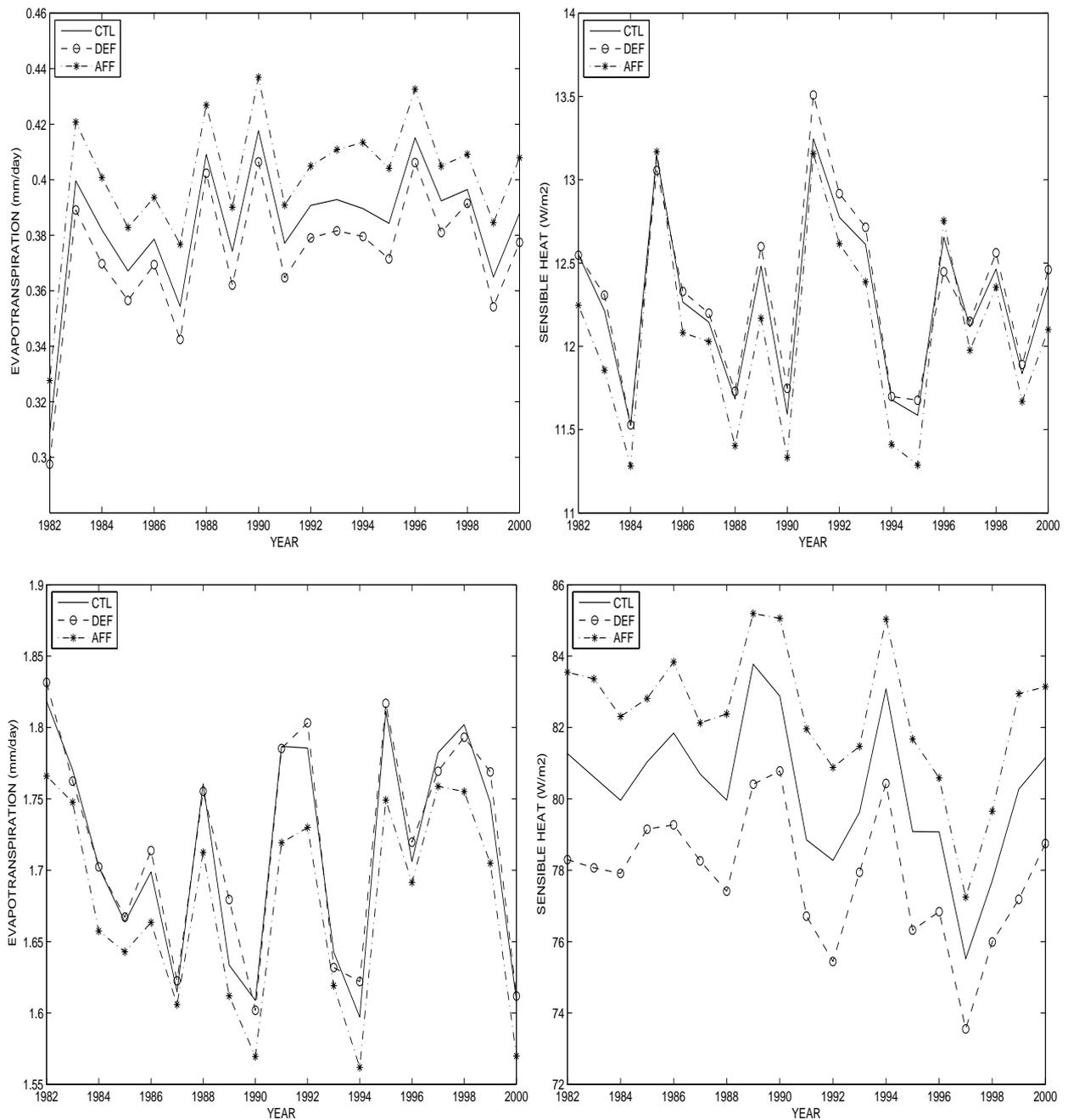


Figure 4.3. Time series of winter evapotranspiration and sensible heat (upper panels) and summer evapotranspiration and sensible heat (lower panels) for the CTL, DEF and AFF experiments.

4.3.2 Spatial differences in seasonally surface climate

We start by discussing the spatial effect of the LCC in the seasonal evolution of the surface climate, in the deforestation case. The most immediate response to vegetation change is found in

the evapotranspiration (and sensible flux) that shows an opposite pattern between winter and summer. In both seasons, the significant vegetation-induced evapotranspiration changes are mainly confined to the Eastern-Northern Europe where the most relevant LCC in the DEF experiment occurs, as confirmed by a two-tailed t -test with a confidence level of 90% (*Figure 4.4a*). As for 2-meter temperature we expected a similar behaviour, with significant differences into the deforested area, instead we found only few statistical significant grid points falling on such area.

In winter, the largest differences in evapotranspiration (*Figure 4.4a*) occur in Scandinavia where the substitution of evergreen needle-leaf forests with crops cause a considerable reduction of LAI and consequently an increase of albedo. Also the Eastern Europe show a reduction of evapotranspiration but in this area the bias is slightly lower than in Scandinavia due to the difference in the vegetation composition. The LCC in these regions implies an increase of surface albedo and a consequent reduction of surface energy available for the soil evaporation.

The winter synoptic systems, taking place in the Atlantic basin and propagating eastward, drive most of the winter atmospheric surface variability over the European continent. The most LCC in DEF experiment are for the Eastern Europe, so they do not affect the synoptic systems entering from the western boundary; consequently only weak changes have been found on winter temperature (*Figure 4.4c*) and precipitation (*Figure 4.4e*) fields.

The largest differences in temperature (*Figure 4.4c*) occur in Scandinavia where the 2-m temperature is warmer than CTL; according to *Sanchez et al.* [2007] this is due to the inverse relationship existing between changes in 2-m temperature and LAI. In fact, the latent heat reduction (*Figure 4.4a*) enhances the sensible heat flux that warms the air above the ground. These anomalies, however, are not statistical significant, except few points falling into deforested area.

The precipitation signal is more complex because of a highly nonlinear feedback with land cover. *Figure 4.4e* shows that areas with a relevant change in precipitation caused by land use change are mainly located along Atlantic coasts of Scandinavia and Iberian Peninsula where in winter the CTL experiment shows the higher precipitations. These differences are due to the reduction of evapotranspiration in the DEF experiment (*Figure 4.4a*); the reduction of water flux to the atmosphere results in less precipitation along the coasts. In the Eastern Europe where occur a deep change from forest to cropland no significant change in precipitation caused by land use change have been found (*Figure 4.4e*). The precipitation over this area could be primarily associated with synoptic scale disturbances entering from the Atlantic boundary during winter time. So, the land-use change does not significantly affect the synoptic disturbances implying few differences of the rainfall field.

In the summer season a positive bias is observed for the evapotranspiration in Eastern-Northern Europe (*Figure 4.4b*). The positive anomaly is mainly due to the effect of the vegetation change onto the local radiative equilibrium; the substitution of mixed forest and evergreen needle-leaf forests with crops reduces the stomatal resistance and the roughness length and in turn increases the evapotranspiration [*Suh and Lee, 2004*].

Also the different soil moisture evolution plays a pivotal role to explain the enhanced summer evapotranspiration and the reversal sign with respect the winter season. In spring, in the northern Europe, the soil is nearly saturated owing to snow melt; moreover the vegetation-induced evaporation decrease during winter (*Figure 4.4a*) leads to a greater amount of water in the soil during the summer season in the DEF experiment (*Figure 4.4b*).

The summer temperature shows a twofold behaviour between Eastern-Northern Europe and central Europe (*Figure 4.4d*). A positive change is observed for the Eastern-northern Europe, where the most relevant LCC of our DEF experiment occurs. The higher evaporation values in Eastern-Northern Europe (*Figure 4.4b*) imply higher water vapour content in the lower troposphere, and higher downward longwave radiation (figure not shown). So, the increased 2m temperature observed in such area it is linked to a positive feedback between vegetation change and local radiation change (discussed later).

On the other hand, the vegetation-induced 2m temperature changes are mostly negative over the southern and Eastern Europe throughout the domain and specifically over the western coast of the Black sea and over the Greek.

It is noteworthy that the area where the maximum anomaly takes place is not affected by vegetation change. This fact is due to the increased evaporation over the Black sea and over the Aegean sea due to wind enhancement, which in turn increases the 2m relative humidity over the western coasts of the Black sea and over Greek. The wind increase could be explained by an overall decrease of the continental roughness length (discussed later).

The evaporation increase over the Eastern Europe, linked to LCC which takes place there, gives explanation of the corresponding rainfall increase (*Figure 4.4f*). In our case, the deforestation causes an increase of the rainfall respect to what reported by *Dumenil-Gates et al. [2001]* mostly over area with strong land-sea contrast. The LCC over the Italian peninsula seems to significantly affect the summer precipitation.

These results of this sensitivity experiment (as simulated by RegCM model) highlight how deforestation in Europe do not change deeply the surface climate: in fact, both temperature and precipitation anomalies are not statistically significant, except few points, with a confidence level greater than 90%. So, in such area, the influence of the baroclinic disturbances on the surface climate is greater than the influence of the vegetation.

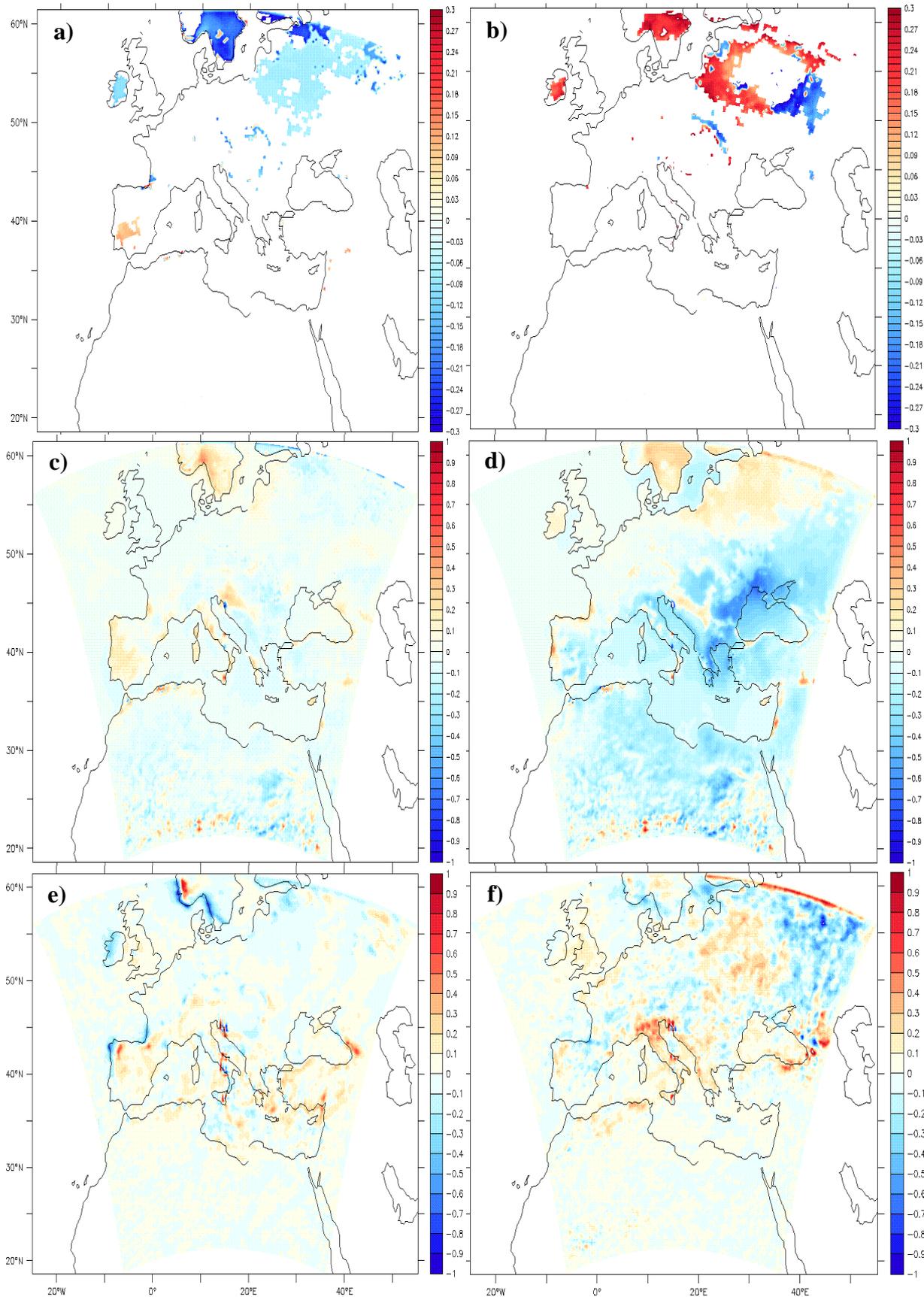


Figure 4.4. DEF minus CTL nineteen-year-averaged anomalies for (a) winter and (b) summer evapotranspiration (mm/day), winter (c) and (d) summer temperature (K), and winter (e) and summer (f) precipitation (mm/day). The anomalies (a) and (b) are statistically significant (t test for 90% confidence level).

Also in the afforestation case the significant vegetation-induced evapotranspiration changes are mainly limited where LCC take place. In the central Europe, where are located the main changes of AFF experiment, the hydrological cycle is characterized by increasing evapotranspiration from April until May (due to the increased solar radiation and transpiration related to the enhanced photosynthetic activity) and by a progressive decreasing values from July to August (due to limited moisture availability and closure of plant stomata) [Heck *et al*, 2001].

During winter, the reduction in the albedo owing to LCC implies a higher net shortwave radiation available at the surface. Due to the soil moisture availability the increased net shortwave energy is used to evaporate (*Figure 4.5a*). The evaporation increases the humidity in the lower troposphere which in turns increases the downward longwave radiation and increases the cloud cover. The later effect dominates over Northern-western Europe where a decrease of the downward short wave is observed.

Downstream the Alps, a strong significant temperature reduction is observed (*Figure 4.5c*). This effect is related to the rainfall reduction upstream the Alps (*Figure 4.5e*). The air parcel impinging the Alps in a northerly/west-northerly flow is lifted following a saturated adiabat. Passing over the mountains, the water vapour is condensed out and removed by precipitation. The air descends on the other side of the mountains following the saturated adiabat, but due to the rainfall removal it will move to a dry adiabat at higher level respect to the ascent. So, the downstream temperature will be higher. In the AFF case the rainfall reduction upstream the Alps will reduce this effect explaining the strong the negative bias in temperature.

The precipitation field (*Figure 4.5e*) features higher precipitation over the Atlantic west-northern Europe coasts. This fact could be explained by the evaporation increase (*Figure 4.5a*) which enhances the water vapour content in the lower troposphere. The ocean SST are higher than the continental surface temperature, so when a warm flow from the sea penetrates onshore it develops instabilities due to the cold land air. In the AFF case, the liquid water content is increased over the coastal area and the large scale precipitation scheme develops rainfall mostly over the coastal area. This precipitation reduces the atmospheric moisture content, in fact the upstream precipitation over the Alps region is reduced (*Figure 4.5e*).

In summer, the albedo reduction allows for more short wave radiation to heat the ground. The soil moisture summer deficit and the vegetation control on the evapotranspiration reduce the evaporation (*Figure 4.5b*) and increase the sensible heat flux. The reduced evaporation decreases the amount of water vapour into the lower troposphere and in turn the greenhouse warming. So, the infrared cooling dominates and the surface temperature is lower respect to the control run (*Figure 4.5d*). The reduced evapotranspiration explains also the negative bias for the summer rainfall (*Figure 4.5f*).

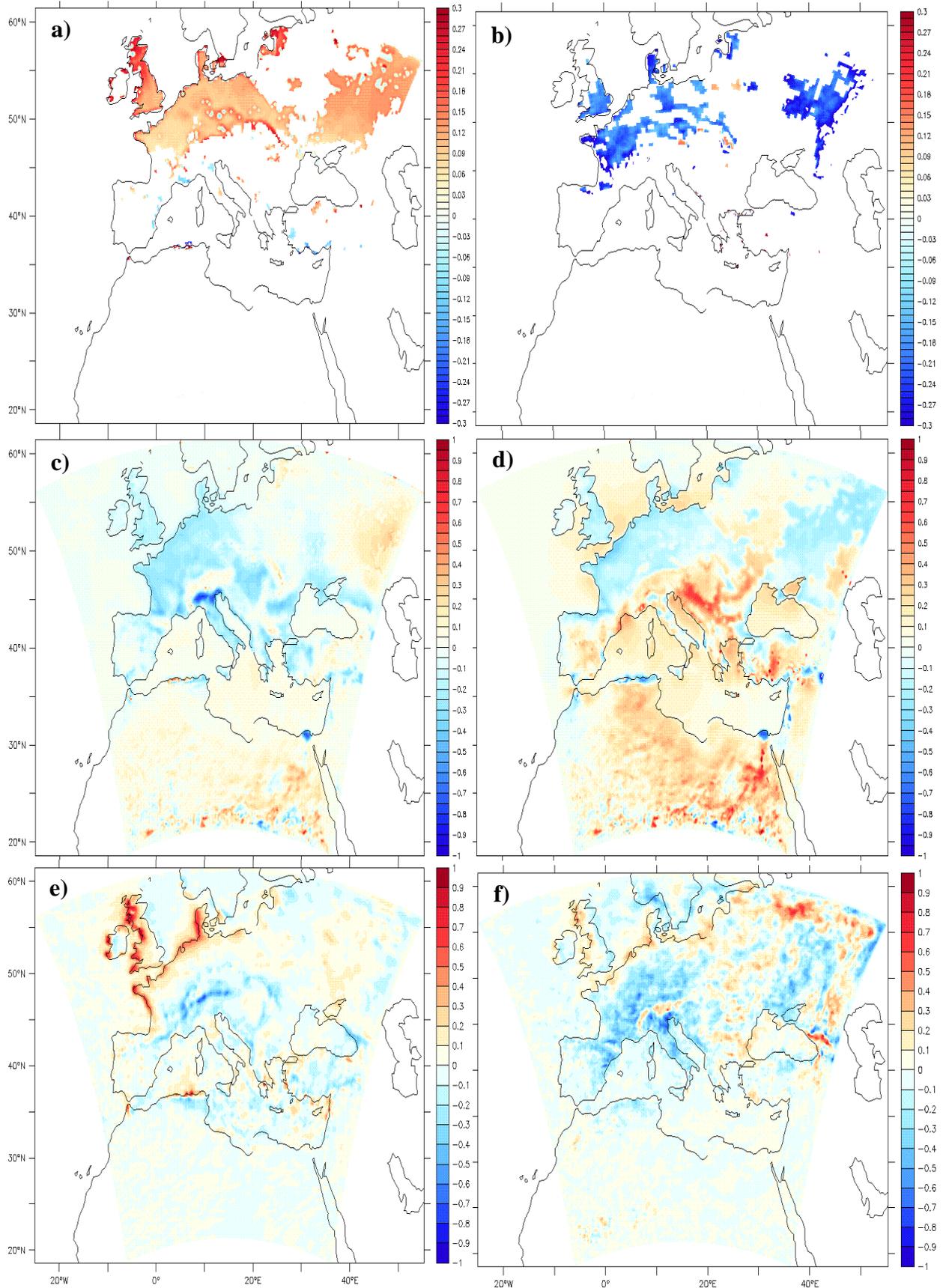


Figure 4.5. AFF minus CTL nineteen-year-averaged differences for (a) winter and (b) summer evapotranspiration (mm/day), winter (c) and (d) summer temperature (K), and winter (e) and summer (f) precipitation (mm/day). The anomalies (a) and (b) are statistically significant (t test for 90% confidence level).

4.3.3 Differences in atmospheric circulation

Now, we verify how the previously analyzed LCC impact on the surface variables, such as temperature, extends to the lower troposphere. We analyze the summer season, which shows significant differences in surface fields between perturbed and control runs. In *Figure 4.6*, we show the wind vector and temperature anomalies at two different levels (925 hPa and 850 hPa) for the Deforestation and Afforestation experiments.

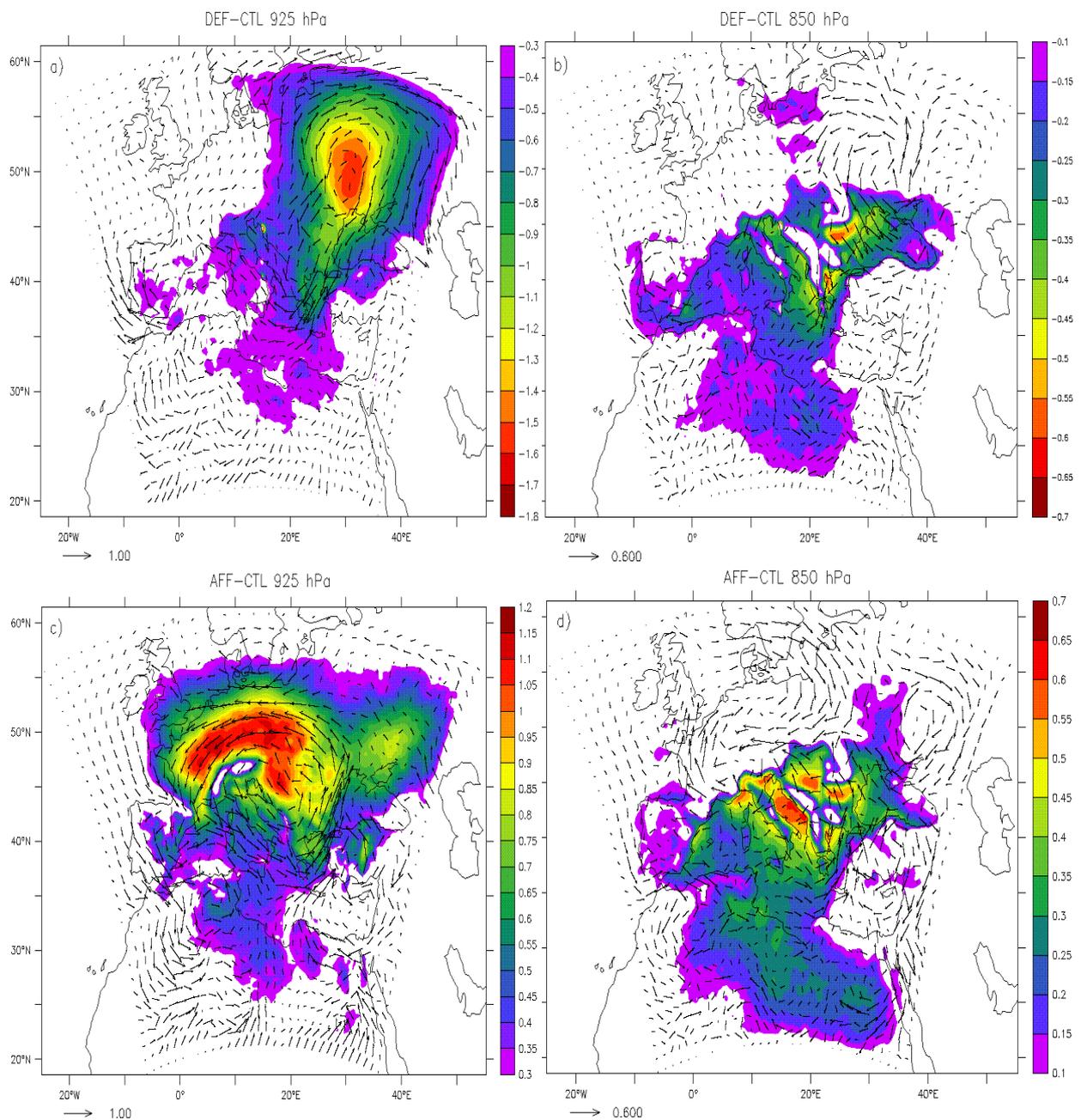


Figure 4.6. Differences (DEF-CTL upper panels, and AFF-CTL lower panels) of horizontal wind vectors (m/s) and temperature (K) at 925 hPa and 850 hPa over the summer season (JJA).

In the Deforestation case, the crop surface albedo lowers the energy availability at the ground level, resulting in a decrease of temperature. Both the stomatal resistance and roughness decreases lead to an increasing of the evapotranspiration. Where the land cover change takes place, the cooling at the surface, increasing the lower level stability, reduces the shallow convection resulting in a surface anti-cyclonic anomaly (*Figure 4.6a*) and in upper cyclonic anomaly (*Figure 4.6b*), where the convections ends.

The increased static stability generates surface anti-cyclonic anomaly by potential vorticity conservation. The surface temperature anomaly is stronger respect to 850 hPa, and the anti-cyclonic circulation advects cool anomaly towards the Black Sea and Eagen Sea, while at 850 hPa the cyclonic circulation advects cool air towards the Black Sea. The near surface temperature anomaly pattern extends beyond the area where the LCC takes place, covering most of the Eastern-Europe and part of the Central Mediterranean up to the Northern Africa.

Considering the Afforestation case, the surface difference flow features a large cyclonic circulation centred over the Adriatic sea. The maximum temperature change is over the western-northern Europe, as for the wind field. Both could be explained by a reduced albedo and an increased roughness length. The mean flow of the control run is mainly westerly over the western-northern Europe, the increased roughness for the Afforestation case tends to rotate clockwise the surface wind, generating a northerly component and reducing the westerly one. This results in the easterly, northern-easterly wind pattern centred at about 50 N (*Figure 4.6c*). An other maximum in the wind field difference is observed over the Greek and Black sea coast, where the Aetesians develop. So, in our simulation the Aetesians are weakened.

The friction increase for the Afforestation case increases the northerly wind component enhancing the wind channelling over the Rhone's valley, and resulting in a stronger wind over the Gulf of Lion respect to the control case.

At 850 hPa (*Figure 4.6d*), the wind differences are weaker, but a significant positive temperature bias is observed over the Central Mediterranean, Balkans and Northern Africa. The excess surface temperature observed over the Western-Central Europe is convected to higher levels and then advected by the mean flow towards the southern Europe and North Africa.

4.3.4 Impact on extreme events

Climate is defined not simply as average temperature and precipitation but also by the type, frequency and intensity of weather events. LCC-induced climate change has the potential to alter the prevalence and severity of extremes such as heat waves, cold waves, storms, floods and

droughts. Though predicting changes in these types of events under a changing climate and under different boundary condition is difficult, understanding vulnerabilities to such changes is a critical part of estimating vulnerabilities and future climate change impacts on human health, society and the environment.

According to the IPCC definition [Solomon *et al*, 2007], the extremes are commonly considered to be the values exceeded 1, 5 and 10% of the time (at one extreme) or 90, 95 and 99% of the time (at the other extreme). The warm nights or hot days are those exceeding the 90th percentile of temperature, while cold nights or days are those falling below the 10th percentile. Heavy precipitation is defined as daily amounts greater than the 95th (or for ‘very heavy’, the 99th) percentile.

In this work, climate extreme events are generally related to unusual values of daily maximum, minimum surface temperatures or precipitation amounts. Hence, changes of these variables between DEF and AFF simulations and control run will be analyzed, considering mean values, selected percentiles (90th for maximum temperature and precipitation) to quantify intensity of extreme processes.

Taking into account the maximum daily temperature, we found an high impact of LCC on the frequency of heat waves. *Figure 4.7* shows the difference in the number of the days during which the temperature is greater than 90th percentile between the two perturbed simulation and the control run.

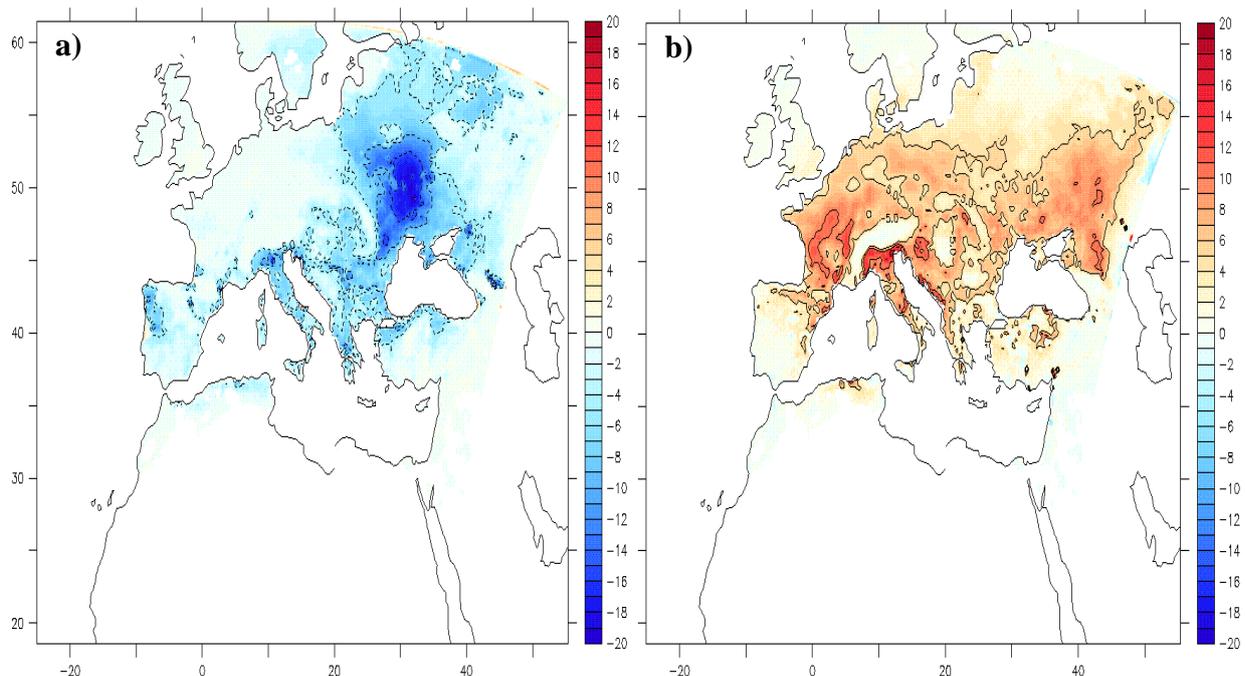


Figure 4.7. Change in the number of summer heat wave days between DEF-CTL (a) and AFF-CTL (b). A heat wave occur when the daily maximum temperature is greater than 90th percentile.

The value of the extreme temperature have been computed from the control run; from this simulations we also computed the number of days with the daily maximum temperature above the threshold of 90th percentile. Hence, this last step have been repeated in the DEF and AFF simulations, and finally we plotted the anomalies in term of days.

Results show a decrease in the number of heat wave events within the DEF experiment (*Figure 4.7a*), with the maximum frequency (around 20 days) centred in the Eastern Europe, where the maximum LCC occurs. This result agree with the general cooling that the deforestation induce and have already been shown above (*Figure 4.6a*). On the other hand, the afforestation generates an increase in the number of summer hot days with respect to the control run. The area where a significant change in frequencies of summer hot waves occur covers all the afforested regions with maximum values up to 18 days in South France and in Italy (downstream the Alps).

These results suggest that the decrease (increase) in the number of summer heat wave days in DEF (AFF) simulations is due to a change of the extremes (i.e. variability and tail of distribution), and not simply to a change in the mean, induced by LCC.

As for precipitation, we found that LCC does not generate summer heavy rain events in both simulations (DEF and AFF).

4.4 DISCUSSIONS AND CONCLUSIONS

Some aspect of land-atmosphere interactions have been investigated by means the regional climate model RegCM3 to identify the sensitivity of the local and regional climate to hypothetical deforestation and afforestation in the Euro-Mediterranean area. The deforestation scenario represents an extreme substitution of forests with crops, especially in the eastern Europe where are confined the main forests. The changes in the afforestation experiment, which lead to a substitution of the crops with forests, are spread over a relatively larger area covering almost all the central Europe.

The simulations reveal a substantial thermodynamically and dynamically consistent impact of vegetation on climate condition in central Europe, particularly during summer. During winter the European atmospheric conditions are determined by translating and developing synoptic-scale disturbances, which are advected into the model domain through its lateral boundaries. As a consequence, vegetation-induced effects cannot locally accumulate to a substantial amplitude, so only weak changes have been found on winter temperature and precipitation fields in both deforestation and afforestation simulations. During summer the weather in southern-central

Europe is frequently characterized by the occurrence of high-pressure system with weak large-scale advection, hence favourable meteorological conditions to analyze the intrinsic atmospheric sensitivity to the land surface. However, also in this season, the only significant changes in surface climate have been found for heat fluxes, while both temperature and precipitation show just few significant points spread over all the domain.

In the deforestation experiment the significant changes in latent heat (and sensible heat) are confined mainly in the eastern Europe, where main LCC take place. In winter the LCC induce a reduction of the evapotranspiration due to both the increased albedo that lowers the soil evaporation, and the lower transpiration of the grass with respect to mixed forest. On the other hand, in summer a reversal sign in latent heat occurs. The enhanced summer evapotranspiration in the deforestation experiment is due to a greater amount of water in the soil layers with respect to the control run as a consequence of the less evapotranspiration during winter.

Also in the afforestation case the significant vegetation-induced evapotranspiration changes are mainly limited where LCC take place. In such case the changes in heat fluxes are opposite those shown in the deforestation experiment. So, the increased latent heat during winter is explained by the reduction in the albedo and, consequently, a greater amount of shortwave radiation available at the surface is used to evaporate. Also in summer, the albedo reduction allows for more short wave radiation to heat the ground. The soil moisture summer deficit and the vegetation control on the evapotranspiration reduce the evaporation.

The t-test results showed how in a given grid point, the mean of temperature in the deforestation (afforestation) experiment is the same to the mean of the control run; in other words, except few spot points, the LCC does not affect the mean of surface temperature (and precipitation). However, significant vegetation-induced changes in temperature have been found in the lower troposphere for summer season. Deforestation results in a cooling of the lower atmosphere, while afforestation acts warming the PBL.

Finally, one of the most important results of this work: as we already stated, in these simulations LCC does not affect the mean of the 2 meter temperature but have a strong impact on the upper tail. This is a quite new result, and for our knowledge never studied before. We found an high impact of LCC on the frequency of heat waves (i.e. the temperature greater than 90th percentile). In particular, with deforestation decrease the frequency of the summer heat waves, while the afforestation increase such frequency. This result agrees with the cooling (warming) that deforestation (afforestation) induce in the PBL.

The results of the cooling (warming) due to the deforestation (afforestation) experiments are in contrast to the impact of vegetation to the GHG; in fact, by afforestation the land surface is able to capture and store in the biomass the atmospheric CO₂, resulting in a less amount of carbon in

the atmosphere and hence a less warming due to the greenhouse effect. For such reason we should consider to develop a new generation of GCM or RCM with an online vegetation able to compute also the CO₂ fluxes and perform new simulation to understand and simulate all the feedbacks between land surface and atmosphere.

We should also note that the results of sensitivity tests depend crucially on the climate conditions of the region where they are performed. It is important highlights that the same LCC may induce different climatic effects over close areas. The effect of LCC is thus dependent on the local environment and climatic conditions.

However, the robustness of the sensitivity is an important result, considering the uncertainties related to several assumptions of this study. For instance, soil type parameters are kept identical in both experiments, even though it is know that erosion is an important process associated with vegetation changes. The atmospheric hydrological cycle is also sensitive to rooting depth. The values used in this study only depends on vegetation type, whereas root characteristics can vary substantially for a single type owing to physical properties of the soil and to adjustments of the relevant species to the local climate and water availability.

Despite its limitations, the setup experiments and the regional climate modeling approach were shown to be a well-suited tool for studying land-atmosphere interactions on a regional scale, a comprehensive level, and physic basis. The regional modeling approach allows the isolation of continental-scale vegetation feedbacks, while global-scale feedbacks are kept under control by conducting the large-scale evolution along the analysis fields at the lateral boundaries of the domain.

Further realism will be added in future to RegCM by the use of interactive vegetation models in which vegetation growth would depend on the simulated climate, with anthropogenic effects included as a disturbance.

APPENDIX

In the Farquhar's model, photosynthesis is calculated as a function of adsorbed photosynthetically active radiation (APAR), temperature, atmospheric CO₂ concentration, day length and canopy conductance [Haxeltine and Prentice, 1996].

➤ PHOTOSYNTHESIS FOR C₃ PATHWAY

As already introduced above, daily gross photosynthesis (A_{gd}) is computed using a standard nonrectangular hyperbola formulation, which gives a gradual transition between two limiting rates. The first one, J_E , describes the response of photosynthesis to APAR:

$$J_E = C1_{C3} * APAR \quad (A.22)$$

where:

$$C1_{C3} = \Phi_C * \Phi_{TC3} * C_{mass} * \alpha_a * \alpha_{C3} * \frac{p_i - \Gamma_*}{p_i + 2\Gamma_*} \quad (A.23)$$

In equation (A.23), Φ_C is a PFT specific parameter which takes into account that the maximum rates of photosynthesis for conifer needles decrease with increasing needle age, C_{mass} (12 g mol⁻¹) is the molar mass of carbon, α_a (Table 3.5) accounts for reduction in PAR utilization efficiencies in natural ecosystem, α_{C3} (Table 3.5) is the intrinsic quantum efficiency for CO₂ uptake and p_i is the internal partial pressure of CO₂, function of the ambient partial pressure of CO₂ (p_a) and the parameter λ (Table 3.5) as described below:

$$p_i = \lambda * p_a \quad (A.24)$$

λ is a key model parameter: many observations have shown that, for C₃ species under non water stressed conditions, stomata responded in a way that maintains a constant ratio of intercellular (p_i) to ambient (p_a) CO₂ partial pressure of 0.6-0.8. Therefore λ is set equal to 0.7 under non water stressed conditions.

The CO₂ compensation point (Γ^*) depend on the partial pressure of oxygen [O_2] and a kinetic parameter (τ) whose temperature dependence is modeled using a Q₁₀ relationship:

$$\Gamma^* = \frac{[O_2]}{2\tau} \quad (A.25)$$

Finally, Φ_{TC3} is a function which models the inhibitory effect of low temperatures on C₃ photosynthesis and it is computed by mean the monthly temperature (T_c) in such way:

$$\Phi_{TC3} = 1 + e^{[0.2*(10-T_c)]^{-1}} \quad (A.26)$$

The second limiting rate used to compute daily non water stressed gross photosynthesis, J_C , describes the Rubisco limited rate of photosynthesis:

$$J_C = C2C3 * V_m \quad (A.27)$$

Where:

$$C2C3 = \frac{p_i - \Gamma^*}{\left[p_i + K_c * \left(1 + \frac{[O_2]}{k_0} \right) \right]} \quad (A.28)$$

and K_c is the Michaelis constant for CO₂ while K_o is the Michaelis constant for O₂.

Now we can write the non water stressed daily gross photosynthesis (A_{gd}) as:

$$A_{gd} = J_E + J_C - \frac{[(J_E + J_C)^2 - 4 * J_C * J_E]^{1/2}}{2\theta} \quad (A.29)$$

As discussed above, the empirical parameter θ (Table 3.5) describes the transition between the two limiting rates J_E and J_C .

Leaf respiration (R_d) depends on the parameter b_{C3} (Table 3.5) and on Rubisco capacity (V_m):

$$R_d = b_{C3} * V_m \quad (A.30)$$

Finally daily net photosynthesis (A_{nd}) is:

$$A_{nd} = A_{gd} - R_d \quad (A.31)$$

The Farquhar's model use an optimization algorithm to predict the values of Rubisco capacity that gives the maximum daily rate of net photosynthesis in the case of non water stressed conditions. This algorithm is based upon hypothesis that the N content and Rubisco activity of leaves vary both seasonally and with canopy position in such way as to maximize net photosynthesis. In LPJ leaf nitrogen content is not modeled explicitly: instead, the optimization is carried out directly on the Rubisco activity of leaves. This optimal value of V_m is calculated by optimizing (A.31) using the constraint $\frac{\partial A_{gd}}{\partial V_m} = 0$, resulting in the following equation:

$$V_m = \left(\frac{1}{b_{C3}} * \frac{C1_{C3}}{C2_{C3}} \right) * [(2\theta - 1) * s - (2\theta s - C2_{C3}) * \sigma] * APAR \quad (A.32)$$

$$s = \frac{24}{d_t} * b_{C3} \quad (A.33)$$

$$\sigma = \sqrt{1 - \left(\frac{C2_{C3} - s}{C2_{C3} - \theta s} \right)} \quad (A.34)$$

where d_t is day length in hours. We can also define total daytime net photosynthesis (A_{dt}) by:

$$A_{dt} = A_{nd} + \left(1 - \frac{d_t}{24} \right) * R_d \quad (A.35)$$

The photosynthetic rate may be related to canopy conductance through the diffusion gradient in CO_2 concentration implied by the difference in CO_2 concentration between the atmosphere and

intercellular air space. Expressing this relationship in terms of the total daytime net photosynthesis (A_{dt}) gives:

$$A_{dt} = \left(\frac{g_c - g_{\min}}{1.6} \right) [c_a (1 - \lambda)] \quad (A.36)$$

where g_c is the average daytime canopy conductance, g_{\min} (Table 3.6) is a PFT specific minimum canopy conductance and c_a is the ambient mole fraction of CO₂. The factor 1.6 accounts for the difference in the diffusion coefficients of CO₂ and water vapor.

Under non-water stressed conditions maximum values of λ are assumed; this allows the calculation of the maximum potential photosynthesis rate and maximum potential canopy conductance (g_{pot}) realizable under non-water stressed conditions.

$$g_{pot} = g_{\min} + 1.6 * \frac{A_{dt}}{c_a (1 - \lambda)} \quad (A.37)$$

Water stress results in lower canopy conductance; in which case the water balance calculation provides a value for the actual water-limited canopy conductance. In order to get the water-limited values of A_{nd} and λ , equation (A.36) is solved simultaneously with equation (A.31) by mean a bisection method.

BIBLIOGRAPHY

- Alessandri A., S. Gualdi, J. Polcher, and A. Navarra (2007). *Effects of Land Surface–Vegetation on the Boreal Summer Surface Climate of a GCM*. *Journal of Climate*, 20(2), 255–278.
- Aster R. C., B. Borchers, and C. H. Thurber (2005). *Parameter Estimation and Inverse Problems*. Elsevier Academic Press, Amsterdam, 301pp.
- Aubinet M., et al. (2000). *Estimates of the annual net carbon and water exchange of forests: The EUROFLUX methodology*. *Advances in Ecological Research*, 30, 113-175.
- Baldocchi D., B. B. Hicks, and T. D. Meyers (1988). *Measuring biosphere-atmosphere exchanges of biologically related gases with micrometeorological methods*. *Ecology* 69, 1331–1340.
- Baldocchi D., D. Falge, et al. (2001). *FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapour, and energy flux densities*. *Bulletin of the American Meteorological Society*, 82 (11), 2415-2434.
- Berthelot M., P. Friedlingstein, P. Ciais P, J-L Dufresne, and P. Monfray (2005). *How uncertainties in future climate change predictions translate into future terrestrial carbon fluxes*. *Global Change Biology*, 11, 959-970.
- Betts A. K., J. H. Ball, C. M. Beljaars, M. J. Miller, and P. A. Viterbo (1996). *The land surface-atmosphere interaction: A review based on observational and global modeling perspectives*. *Journal of Geophysical Research*, 101, 7209-7225.
- Betts R. A., P. M. Cox, S. E. Lee, and F. I. Woodward (1997). *Contrasting physiological and structural vegetation feedbacks in climate change simulations*. *Nature*, 387, 796–799.
- Betts R. A. (2001). *Biogeophysical impacts of land use on present-day climate: near-surface temperature change and radiative forcing*. *Atmospheric Science Letters*, 1.
- Bolin B., P. Ciais, W. P. Cramer, P. G. Jarvis, H. S. Kheshgi, C. A. Nobre, S. M. Semenov, W. L. Steffen, S. Linder, and F. Joos (2000). *Land use, land-use change and forestry*. Cambridge University Press, Cambridge.
- Bonan G. B, K. Oleson, M. Vertenstein, S. Levis, X. Zeng, Y. Dai, R. E. Dickinson, and Y. Zong-Liang (2002). *The Land Surface Climatology of the Community Land Model Coupled to the NCAR Community Climate Model*. *Journal of Climate*, Volume 15, Issue 22, 1307–1326.
- Bounoua L., G. J. Collatz, C. J. Tucker, P. J. Sellers, D. A. Randall, et al. (1999). *Interactions between vegetation and climate: Radiative and physiological effects of doubled atmospheric CO₂*. *Journal of Climate*, 12, 309–324.

-
- Braswell B. H., W. J. Sacks, E. Linder, and D. S. Schimel (2005). *Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations*. *Global Change Biology*, 11(2), 335–355.
 - Bratley P., B. L. Fox, L. E. Schrage (1987). *A Guide to Simulation*. Springer, New York.
 - Burke I. C., D. S. Schimel, C. M. Yonker, W. J. Parton, L. A. Joyce, and W. K. Lauenroth (1991). *Regional modeling of grassland biogeochemistry using GIS*. *Landscape Ecology* 4, 45–54.
 - Chapin F. S., P. A. Matson, and H. A. Mooney (2002). *Principles of terrestrial ecosystem ecology*. Springer-Verlag New York.
 - Chester R. (2000). *Dissolved gases in sea water. Marine Geochemistry*, 2nd ed., Blackwell Science, Oxford, UK.
 - Ciais P., P. P. Tans, M. Trolier, J. W. White, and R. J. Francey (1995). *A large northern hemisphere terrestrial CO₂ sink indicated by the C₁₃/C₁₂ ratio of atmospheric CO₂*. *Science*, 269 (5227), 1098–1102.
 - Ciais P., A. S. Denning, P. P. Tans, et al. (1997). *A three-dimensional synthesis study of δ¹⁸O in atmospheric CO₂: 1. Surface fluxes*. *Journal of Geophysical Research*, Vol. 102(D5), 5857–5872.
 - Ciais P., et al. (2005). *Europe-wide reduction in primary productivity caused by the heat and drought in 2003*. *Nature*, 437, 22.
 - Collatz G. J., J. T. Ball, C. Grivet, and J. A. Berry (1991). *Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer*. *Agricultural and forest meteorology*, 54, 107–136.
 - Collatz G. J., M. Ribas-Carbo, and J. A. Berry (1992). *A coupled photosynthesis stomatal conductance model for leaves of C₄ plants*. *Australian Journal of Plant Physiology*, 19, 519–538.
 - Conover W. J., and R. L. Iman (1981). *Rank transformation as a bridge between parametric and non-parametric statistics*. *American Statistician*, 35, 124–129.
 - Cox P. (2001). *The terrestrial carbon cycle model TRIFFID*. Hadley Centre, Oxford, 4 pp.
 - Cox P. M., R. A. Betts, M. Collins, P. P. Harris, C. Huntingford, and C. D. Jones (2004). *Amazon dieback under climate-carbon cycle projections for the 21st century*. *Theoretical Applied Climatology*, 78, 137–156.
 - Cramer W., et al. (1999). *Comparing global models of terrestrial net primary productivity (NPP): Overview and key results*. *Global Change Biology*, 5, Suppl. 1, 1–15.
 - 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 Biology*, 7, 357–373.
 - Cramer W., A. Bondeau, S. Schaphoff, W. Lucht, B. Smith, and S. Sitch (2004). *Tropical forests and the global carbon Cycle: Impacts of atmospheric carbon dioxide, climate change and rate of*

deforestation. Philosophical Transactions of the Royal Society of London-Series B, 359, 331-343.

- Denman K. L., et al. (2007). *Climate Change 2007: The Physical Science Basis*. Working Group 1 Contribution to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC).
- Dickinson R. E. (1983). *Land surface processes and climate surface albedos and energy balance*. Advanced Geophysics, 25, 305-353.
- Dickinson R. E., A. Henderson-Sellers, P. J. Kennedy, and M. F. Wilson (1986). *Biosphere–Atmosphere Transfer Scheme (BATS) for the NCAR Community Climate Model*. NCAR Tech. Note NCAR/TN275+STR, 69 pp. [Available from National Centre for Atmospheric Research, Boulder, CO 80307].
- Diemont W. H. (1994). *Effects of removal of organic matter on the productivity of heathlands*. Journal of Vegetation Science, 5, 409-414.
- Dore S., G. J. Hymus, D. P. Johnson, C. R. Hinkle, R. Valentini, and B. G. Drake (2003). *Cross validation of open-top chamber and eddy covariance measurements of ecosystem CO₂ exchange in a Florida scrub-oak ecosystem*. Global Change Biology, 9(1), 84–95.
- Dumenil Gates L., and Ließ (2001). *Impacts of deforestation and afforestation in the Mediterranean region as simulated by the MPI atmospheric GCM*. Global and Planetary Change, 30, 309-328.
- Entekhabi D. (1995). *Recent advanced in land-atmosphere interaction research*. Reviews Of Geophysics, 33, S1, 995–1004.
- Ettema J., and P. Viterbo (2004). *ECMWF soil moisture data assimilation*. ECMWF/ELDAS Workshop on Land Surface Assimilation. 8-11/11/2004, pg 97.
- Evensen G. (2006). *Data Assimilation*. Springer.
- Fahey B., and R. B Jackson (1997). *Hydrological impacts of converting native forests and grasslands to pine plantations, South Island, New Zealand*. Agric. Forest Meteorology, 84, 69–82.
- Falge E., et al. (2001). *Gap filling strategies for defensible annual sums of net ecosystem exchange*. Agricultural and Forest Meteorology, 107(1), 43-69.
- FAO (2001). *Global forest resources assessment 2000: Main report*. Food and Agriculture Organisation of the United Nations, Rome, 479 pp.
- Farquhar G. D., S. Von Caemmerer, and J. A. Berry (1980). *A biochemical model of photosynthetic CO₂ assimilation in leaves of C₃ species*. Planta, 149, 78–90.
- Farquhar G. D., J. Lloyd, J. A. Taylor, et al. (1993). *Vegetation effects on the isotope composition of oxygen in atmospheric CO₂*. Nature, 363, 439-442.

-
- Ferranti L., and P. Viterbo (2006). *The European summer of 2003: Sensitivity to soil water initial conditions*. Journal of Climate, 19, 3659–3680.
 - Fischlin A., G. F. Midgley, J. Price, et al. (2007). *Climate Change 2007: Impacts, Vulnerability and Adaptation*. Chapter 4 to Working Group 2 Contribution to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC), pp 211-277.
 - Foley J., J. E. Kutzbach, M. T. Coe, et al. (1994). *Feedbacks between climate and boreal forests during the Holocene epoch*. Nature, 371, 52-54.
 - Foley J. (1995). *An equilibrium model of the terrestrial carbon budget*. Tellus, 47B, 310–319.
 - Foley J., I. C. Prentice, N. Ramankutty, S. Levis, D. Pollard, S. Sitch, and A. Haxeltine (1996). *An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics*. Global Biogeochemical Cycles, 10, 603–628.
 - Fouillet A., et al. (2006). *Excess mortality related to the August 2003 heat wave in France*. International archives of occupational and environmental health, 80, 16-24.
 - Friedlingstein P., P. Cox, et al. (2006). *Climate–Carbon Cycle Feedback Analysis: Results from the C4MIP Model Intercomparison*. Journal of Climate, 19, 3337-3353.
 - Friend A. D., et al. (2007). *FLUXNET and modeling the global carbon cycle*. Global Change Biology 13(3), 610–633.
 - Gaertner M. A., O. B. Christensen, J. A. Prego, J. Polcher, C. Gallardo, and M. Castro (2001). *The impacts of deforestation on the hydrological cycle in the western Mediterranean: an ensemble study with two regional models*. Climate Dynamics, 17, 857-873.
 - Gao X., Y. Luo, W. Lin, Z. Zhao, and F. Giorgi (2003). *Simulation of effects of land use change on climate in China by a regional climate model*. Advanced in Atmospheric Sciences, 20, 4, 583-592.
 - Gauthier P. (1992). *Chaos and quadri-dimensional data assimilation: a study based on the Lorenz model*. Tellus 44A, 2–17.
 - Gelman A., J. B. Carlin., H. S. Stern, and D. B. Rubin (1995). *Bayesian Data Analysis*. Chapman and Hall/CRC, New York.
 - Gerber S., F. Joos, and I. C. Prentice (2004). *Sensitivity of a dynamic global vegetation model to climate and atmospheric CO₂*. Global Change Biology, 10, 1223–1239.
 - Giorgi F. (1990). *Simulation of regional climate using a limited area model nested in a general circulation model*. Journal of Climate, 3, 941-963.
 - Giorgi, F., and L.O. Mearns (1991). *Approaches to regional climate change simulation: A review*. Review of Geophysics, 29, 191-216.
 - Giorgi F, M. R. Marinucci, and G. T. Bates (1993a). *Development of a second generation regional climate model (RegCM2). Part I. Boundary-layer and radiative transfer processes*. Mon. Weather Rev. 121,794-813.

-
- Giorgi F, M. R. Marinucci, G. T. Bates, and G. De Canio (1993b). *Development of a second generation regional climate model(RegCM2). Part II. Convective processes and assimilation of lateral boundary conditions*. Mon. Weather Rev., 121,814-832.
 - Giorgi F, L. O. Mearns (1999). *Introduction to special section: regional climate modeling revisited*. J. Geophys. Res., 104, 335-352.
 - Gitay H., S. Brown, W. E. Easterling, et al. (2001). *Climate change 2001: Impacts, Adaptation, and Vulnerability; Contribution of working group II to the 3rd assessment report of the IPCC*. Cambridge University Press, London & New York pp. 235-342.
 - Gorham E. (1991). *Northern peatlands: role in the carbon cycle and probable responses to climatic warming*. Ecological Applications, 1, 181-195.
 - Goudriaan J., H. Shugart, H. Bugmann, W. Cramer, et al. (1999). *The terrestrial biosphere and global change. Implications for natural and managed ecosystems*. Cambridge University Press.
 - Grace J. (2004). *Understanding and managing the global carbon cycle*. Journal of Ecology, 92, 189- 202.
 - Gruber N., P. Friedlingstein, C. B. Field, R. Valentini, M. Heinmann, J. E. Richey, P. R. Lankao, E. D. Schulze, and C. T. Chen (2004). *The global carbon cycle: Integrating humans, climate and the natural world*. Island Press, Washington.
 - Hargreaves J. C., and J. D. Annan (2002). *Assimilation of paleo-data in a simple Earth system model*. Climate Dynamics, 19, 371–381.
 - Hartmann D. L. (1994). *Global Physical Climatology*. International Geophysics Series, Ed. R. Dmowska and J.R. Holton, Academic Press, Vol. 56, 411 pp.
 - Haxeltine A., and I. C. Prentice (1996). *BIOME 3: an equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability and competition among plant functional types*. Global Biogeochemical Cycles, 10(4), 693–703.
 - Heck P., D. Luthi, H. Wernli, and C. Schar (2001). *Climate impacts of European-scale anthropogenic vegetation changes: a sensitivity study using a regional climate model*. Journal of Geophysical Research, 106, 7817– 7835.
 - Hitz S., and J. B. Smith (2004). *Estimating global impacts from climate change*. Global Environmental Change, 14, 201-218.
 - Hollinger D. Y., and A. D. Richardson (2005). *Uncertainty in eddy covariance measurements and its application to physiological models*. Tree Physiology, 25, 873–885.
 - Houghton J. T., L. G. Meira Filho, B. A. Callander, N. Harris, A. Kattenberg, and K. Maskell (1995). *Climate Change 1995. The Science of Climate Change*. Cambridge University Press, 572 pp.
 - Houghton J. T., D. J. Ding, M. Griggs, P. J. Noguera, Van Der Linden, X. Dai, K. Maskell, and C. A. Johnson (2001). *Climate Change 2001: The Scientific Basis*. Cambridge University Press, New York.

-
- Houghton R. A. (2003). *Revised estimates of the annual flux of carbon to the atmosphere changes in land use and land management 1950-2000*. Tellus B, 55, 378-390.
 - Houghton R. A. (2002). *Magnitude, distribution and causes of terrestrial carbon sinks and some implications for policy*. Climate Policy, 69, 1-18.
 - House J. I., I. C. Prentice, N. Ramankutty, R. A. Houghton, and M. Heimann (2003). *Reconciling apparent inconsistencies in estimates of terrestrial CO₂ sources and sinks*. Tellus B, 55 (2), 345-363.
 - Indermühle A., et al. (1999). *Holocene carbon cycle dynamics based on CO₂ trapped in ice at Taylor Dome, Antarctica*. Nature 398 (6723), 121–126.
 - Keddy P. A. (2000). *Wetland ecology: Principles and conservation*. Cambridge University Press, Cambridge, UK., 614 p.
 - Keeling C. D., and T. P. Whorf (2005). *Atmospheric CO₂ records from sites in the SIO air sampling network*. In Trends: A Compendium of Data on Global Change. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A.
 - Kicklighter D. W., M. Bruno, S. Dönges, G. Esser, M. Heimann, J. Helfrich, et al. (1999). *A first-order analysis of the potential role of CO₂ fertilization to affect the global carbon budget: a comparison of four terrestrial biosphere models*. Tellus B, Vol. 51, issue 2, p.343.
 - Knorr W. (2000). *Annual and interannual CO₂ exchanges of the terrestrial biosphere: process-based simulations and uncertainties*. Global Ecology and Biogeography, 9, 225–252.
 - Knorr W., and M. Heimann (2001). *Uncertainties in global terrestrial biosphere modeling. Part I: a comprehensive sensitivity analysis with a new photosynthesis and energy balance scheme*. Global Biogeochemical Cycles, 15, 207–225.
 - Knorr W., and J. Kattge (2005). *Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling*. Global Biogeochemical Cycles, 15, 207–225.
 - Kobler A., T. Cunder, and J. Pirnat (2005). *Modeling spontaneous afforestation in Postojna area, Slovenia*. Journal for Nature Conservation 13, 2/3, 127-135.
 - Körner C. (2003). *Slow in, Rapid out-Carbon Flux Studies and Kyoto Targets*. Science, 300, 1242-1243.
 - Koster et al., (2004). *Regions of strong coupling between soil moisture and precipitation*. Science, 305, 1138-1140.
 - Krinner G., et al. (2005). *A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system*. Global Biogeochemical Cycles, 19.
 - Kruijt B., et al. (2004). *The robustness of eddy correlation fluxes for Amazon rain forest conditions*. Ecological Applications, 14, 101–113.

-
- Kucharik C., et al. (2000). *Testing the performance of a dynamic global ecosystem model: Water balance, carbon balance, and vegetation structure*. *Global Biogeochemical Cycles*, 14, 795–825.
 - Lambin E. F., H. J. Geist, and E. Lepers (2003). *Dynamics of land-use and land-cover change in tropical regions*. *Annual Review of Environment and Resources*, 28, 205-241.
 - Levis S., J. A. Foley, and D. Pollard (2000). *Large-scale vegetation feedbacks on a doubled CO₂ climate*. *Journal of Climate*, 13, 1313–1325.
 - Lloyd J., and J. A. Taylor (1994). *On the temperature dependence of soil respiration*. *Functional Ecology*, 8, 315–323.
 - Lloyd J., and G. D. Farquhar (1996). *The CO₂ dependence of photosynthesis, plant growth responses to elevated atmospheric CO₂ concentrations and their interaction with soil nutrient status. I. General principles and forest ecosystems*. *Functional Ecology*, 10, 4-32.
 - Loveland T. R., J. W. Merchant, D. O. Ohlen, and J. F. Brown (1991). *Development of a Land Cover Characteristics Database for the Conterminous US*. *Photogrammetric Engineering and Remote Sensing*, 57(11), 1453-1463.
 - Loveland T. R., B. C. Reed, et al. (2000). *Development of a Global Land Cover Characteristics Database and IGBP DISCover from 1-km AVHRR Data*. *International Journal of Remote Sensing*, 21, 6/7, 303-330.
 - Lucht W., S. Schaphoff, T. Erbrecht, U. Heyder, and W. Cramer (2006). *Terrestrial vegetation redistribution and carbon balance under climate change*. *Carbon Balance and Management*, 1:6.
 - Magnani F., S. Leonardi, R. Tognetti, J. Grace, and M. Borghetti (1998). *Modeling the surface conductance of a broad-leaf canopy: Effects of a partial decoupling from the atmosphere*. *Plant, Cell and Environment*, 21, 867-879.
 - Mary B., S. Recous, and D. Robin (1998). *A model for calculating nitrogen fluxes in soil using ¹⁵N tracing*. *Soil Biology & Biochemistry* 30, 1963–1979.
 - Matear R. J. (1995). *Parameter optimization and analysis of ecosystem models using simulated annealing: a case study at Station P*. *Journal of Marine Research*, 53, 571-607.
 - McGuire A. D., et al. (2001). *Carbon balance of the terrestrial biosphere in the twentieth century: analyses of CO₂, climate and land use effects with four process-based ecosystem models*. *Global Biogeochemical Cycles*, 15, 183–206.
 - McMurtrie R. E., and Y.-P. Wang (1993). *Mathematical models of the photosynthetic response of tree stands to rising CO₂ concentrations and temperatures*. *Plant Cell Environment*, 16, 1-13.
 - Meyer W. B., and B. L. Turner (1998). *Changes in land use and land cover: a global perspective*. Cambridge University Press, Cambridge, 537 pp.
 - Metropolis N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). *Equation of state calculations by fast computing machines*. *Journal of Chemical Physics*, 21, 1087–1092.

-
- Miller R. N., M. Ghil, and F. Gauthiez (1994). *Advanced data assimilation in strongly nonlinear dynamical systems*. *Journal of Atmospheric Science*, 51(8), 1037–1056.
 - Mitchell T. D., and P. D. Jones (2005). *An improved method of constructing a database of monthly climate observations and associated high resolution grids*. *Int. J. Climatology*, 25(6), 693–712.
 - Moncrieff J. B., Y. Malhi, and R. Leuning (1996). *The propagation of errors in long-term measurements of land-atmosphere fluxes of carbon and water*. *Global Change Biology*, 2, 231–240.
 - Monteith J. L. (1995). *Accommodation between transpiring vegetation and the convective boundary layer*. *Journal of Hydrology*, 166, 251–263.
 - Morales P., et al. (2005). *Comparing and evaluating process-based ecosystem model predictions of carbon and water fluxes in major European forest biomes*. *Global Change Biology*, 11(12), 2211-2233.
 - Morales P., T. Hickler, D. P. Rowell, B. Smith, and M. T. Sykes (2007). *Changes in European ecosystem productivity and carbon balance driven by regional climate model output*. *Global Change Biology*, 13, 108-122.
 - Mosegaard K., and A. Tarantola (1995). *Monte Carlo sampling of solutions to inverse problems*. *Journal of Geophysical Research*, 100, 12431–12447.
 - Mosegaard K. (1998). *Resolution analysis of general inverse problems through inverse Monte Carlo sampling*. *Inverse Problems*, 14, 405-426.
 - Mosegaard K. (2006). *Monte Carlo analysis of inverse problems*. Ph.D. thesis, University of Copenhagen, København, 17 March.
 - Muller C., T. Rütting, J. Kattge, R. L. Laughlin, and R. J. Stevens (2007). *Estimation of parameters in complex ¹⁵N tracing models by Monte Carlo sampling*. *Soil Biology & Biochemistry*, 39(3), 715-726.
 - Nemani R. R., C. D. Keeling, et al. (2003). *Climate-Driven Increases in Global Terrestrial Net Primary Production from 1982 to 1999*. *Science*, 300, 1560-1563.
 - Noretto M. D., E. G. Jobbagy, and J. M. Paruelo (2005). *Land-use change and water losses: the case of grassland afforestation across a soil textural gradient in central Argentina*. *Global Change Biology*, 11, 1101–1117.
 - Notaro M., Zhengyu Liu, R. Gallimore, J. S. Vavrus, and J. E. Kutzbach (2004). *Simulated and observed preindustrial to modern vegetation and climate changes*. *Journal Of Climate*, 18, 3650-3671.
 - Ojima D. S., B. O. Dirks, E. P. Glenn, C. E. Owensby, and J. O. Scurlock (1993). *Assessment of C budget for grasslands and drylands of the world*. *Water Air and Soil Pollution*, 70, 95-109.

-
- Pal J. S., E. Small, and E. Eltahir (2000). *Simulation of regional-scale water and energy budgets: representation of subgrid cloud and precipitation processes within RegCM*. J. Geophys. Res. 105, 579–594.
 - Pal J. S., et al. (2007). *The ICTP RegCM3 and RegCNET: regional climate modeling for the developing world*. B. Am. Meteorol. Soc., in press, 2007.
 - Pan Y., A. McGuire, J. M. Melillo, D. W. Kicklighter, S. Sitch, and I. C. Prentice (2002). *A biogeochemistry-based dynamic vegetation model and its application along a moisture gradient in the continental United States*. Journal of Vegetation Science, 13(3), 369-382.
 - Papale D., M. Reichstein, E. Canfora, M. Aubinet, C. Bernhofer, B. Longdoz, W. Kutsch, S. Rambal, R. Valentini, T. Vesala, and D. Yakir (2006). *Towards a more harmonized processing of eddy covariance CO₂ fluxes: algorithms and uncertainty estimation*. Biogeosciences Discussion, 3, 961–992.
 - Parton W. J., D. S. Ojima, and D. Schimel (1994). *Environmental change in grasslands: Assessment using models*. Climatic Change, 28, 111-141.
 - Pielke R. A., R. Avissar, M. Raupach, A. J. Dolman, X. Zeng, and S. Denning (1998). *Interactions between the atmosphere and terrestrial ecosystems: influences on weather and climate*. Global Change Biology, 4, 461-475.
 - Pielke R. A., J. Adegoke, A. Beltran-Przekurat, C. A. Hiemstra, J. Lin, U. S. Nair, D. Niyogi and T. E. Nobis (2007). *An overview of regional land-use and land-cover impacts on rainfall*. Tellus, 59B, 587–601.
 - Pitman A. J. (2003). *The evolution of, and revolution in, land surface schemes designed for climate models*. Int. J. Climatol., 23, 479–510.
 - Potter C. S., S. Klooster, and V. Brooks (1999). *Interannual variability in terrestrial net primary production: Exploration of trends and controls on regional to global scales*. Ecosystems, 2, 36-48.
 - Prentice I. C., M. Heimann, and S. Sitch (2000). *The carbon balance of the terrestrial biosphere: ecosystem models and atmospheric observations*. Ecological Applications, 10(6), 1553-1573.
 - Prentice I. C., G. C. Farquhar, M. J. R. Fasham, et al. (2001). *Climate Change 2001. The Scientific Basis*. Cambridge University Press, Chapter 3.
 - Press W. H., S. A. Teukolsky, V. T. Vetterling, and B. P. Flannery (1992). *Numerical Recipes in C. The Art of Scientific Computing*. Cambridge University Press.
 - Quinn G. P., and M. J. Keough (2002). *Experimental Design and Data Analysis for Biologists*. Cambridge University Press, Cambridge, 537pp.
 - Reale O., and J. Shukla (2000). *Modeling the effects of vegetation on Mediterranean climate during the Roman Classical Period: Part II. Model simulation*. Global and Planetary Change, 25, 185-214.

-
- Reichstein M., et al. (2005). *On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm*. *Global Change Biology*, 11, 1424-1439.
 - Reichstein M., et al. (2007). *Reduction of ecosystem productivity and respiration during the European summer 2003 climate anomaly: a joint flux tower, remote sensing and modeling analysis*. *Global Change Biology*, 13, 634-651.
 - Reynolds R.W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang (2002). *An improved in situ and satellite SST analysis for climate*. *J. Climate*, 15, 1609-1625.
 - Sabine C. L., M. Heimann, P. Artaxo, D. C. Bakker, C. T. Chen, C. B.G. Field, N. Gruber, C. L. Quere, R. G. Prinn, J. E. Richey, P. R. Lankao, J. A. Sathaye, and R. Valentini (2004). *The global carbon cycle: Integrating humans, climate and natural world*. Island Press, Washington. pp. 17-44.
 - Sacks W. J., D. S. Schimel, R. K. Monson, and B. H. Braswell (2006). *Model-data synthesis of diurnal and seasonal CO₂ fluxes at Niwot Ridge, Colorado*. *Global Change Biology*, 12, 240-259.
 - Sanchez E., M. A. Gaertner, C. Gallardo, E. Padorno, A. Arribas, and M. Castro (2007). *Impacts of a change in vegetation description on simulated European summer present-day and future climates*. *Climate Dynamics*, 29, 319-332.
 - Santaren D., P. Peylin, N. Viovy, and P. Ciais (2007). *Optimizing a process-based ecosystem model with eddy-covariance flux measurements: A pine forest in southern France*. *Global Biogeochemical Cycles*, 21.
 - Schaphoff S., W. Lucht, D. Gerten, S. Sitch, W. Cramer, and I. C. Prentice (2006). *Terrestrial biosphere carbon storage under alternative climate projections*. *Climatic Change*, 74, 97-122.
 - Schär C, P. L. Vidale, D. Lüthi, C. Frei, C. Häberli, M. A. Liniger, and C. Appenzeller (2004). *The role of increasing temperature variability in European summer heat waves*. *Nature*, 427, 332-336.
 - Schmid H. P. (1997). *Experimental design for flux measurements: matching scales of observations and fluxes*. *Agricultural and forest meteorology* 87, 179-200.
 - Scholes R. J., and I. R. Noble (2001). *Storing Carbon on Land*. *Science*, 294, 1012-1013.
 - Sellers P. J., Y. Mintz, Y. C. Sud, and A. Dalcher (1986). *A simple biosphere model (SiB) for use within general circulation models*. *Journal of Atmospheric Sciences*, 43, 505-531.
 - Sellers P. J., F. G. Hall, H. Margolis, et al. (1995). *The boreal ecosystem-atmosphere study (BOREAS): an overview and early results from the 1994 field year*. *Bulletin of the American Meteorological Society*, 76, 1549-1577.
 - Sellers P. J., L. Bounoua, G. J. Collatz, D. A. Randall, D. A. Dazlich, S. O. Los, et al (1996). *Comparison of radiative and physiological effects of doubled atmospheric CO₂ on climate*. *Science*, 271, 1402-1406.

-
- Seneviratne S., D. Luthi, M. Litschi, and C. Schar (2006). *Land-Atmosphere coupling and climate change in Europe*. Science, 443, 205-209.
 - Silver W. L. (1998). *The potential effects of elevated CO₂ and climate change on tropical forest soils and biogeochemical cycling*. Climatic Change, 39, 337-361.
 - Sitch S., B. Smith, I. C. Prentice, A. Arneth, A. Boundeau, W. Cramer, et al. (2003). *Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model*. Global Change Biology, 9, 161-185.
 - Smith T. M., H. H. Shugart, and F. I. Woodward (1997). *Plant Functional Types: Their Relevance to Ecosystem Properties and Global Change*. Cambridge University Press, 369 pp.
 - Smith B., I. C. Prentice, and M. T. Sykes (2001). *Representation of vegetation dynamics in the modeling of terrestrial ecosystems: comparing two contrasting approaches within European climate space*. Global Ecology and Biogeography, 10, 621-637.
 - Solomon S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (2007). *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
 - Suh, M.-S., and D.-K. Lee (2004). *Impacts of land use/cover changes on surface climate over east Asia for extreme climate cases using RegCM2*. J. Geophys. Res., 109.
 - Tarantola A. (2005). *Inverse problem theory and methods for model parameter estimation*. Society for Industrial & Applied Mathematics, Philadelphia.
 - Trudinger C. M., et al. (2007). *OptIC project: An intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models*. Journal of Geophysical Research, 112, G02027.
 - Uppala S. M., P. W. Kållberg, et al. (2005). *The ERA-40 re-analysis*. Quart. J. R. Meteorol. Soc., 131, 2961-3012.
 - Valentini R., G. Matteucci, A. J. Dolman, E. D. Schulze, C. Rebmann, E. J. Moors, A. Granier, et al. (2000). *Respiration as the main determinant of carbon balance in European forests*. Nature, 404, 861-865.
 - Valentini R., (2003). *Fluxes of Carbon, Water and Energy of European Forests*. Ecological Studies, Vol. 163
 - Van Oijen M., J. Rougier, and R. Smith (2005). *Bayesian calibration of process-based forest models: bridging the gap between models and data*. Tree Physiology, 25, 915-927.
 - Vautard R., et al. (2007). *Summertime European heat and drought waves induced by wintertime Mediterranean rainfall deficit*. Geophys. Res. Lett., 34, L07711.
 - Wang Y.-P., R. Leuning, A. H. Cleugh, and P. A. Coppin (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(5), 495-510.

-
- Wang, Y.-P., D. Baldocchi, R. Leuning, E. Falge, and T. Vesala (2007). *Estimating parameters in a land-surface model by applying nonlinear inversion to eddy covariance flux measurements from eight FLUXNET sites*. *Global Change Biology* 13, 652-670.
 - Waring R. H., J. J Landsberg, and M. Williams (1998). *Net primary production of forests: A constant fraction of gross primary production?* *Tree Physiology*, 18, 129-134.
 - Watson R. T., I. R. Noble, B. Bolin, N. H. Ravindranth, D. Verado, and D. J. Dokken (2000). *IPCC Special report on Land Use, Land-Use Change, and Forestry*. Cambridge University Press, Cambridge, 377 pp.
 - White M. A., P. E. Thornton, S. W. Running, et al. (2000). *Parameterization and sensitivity analysis of the BIOME-BGC terrestrial ecosystem model: net primary production controls*. *Earth Interactions*, 4, 1-85.
 - Williams M., P. A. Schwarz, B. E. Law, J. Irvine, and M. R. Kurpius (2005). *An improved analysis of forest carbon dynamics using data assimilation*. *Global Change Biology*, 11, 89-105.
 - Wilson K., et al., (2002). *Energy balance closure at FLUXNET sites*. *Agricultural and Forest Meteorology*, 113, 223-243.
 - Woodward F. I., and W. Cramer (1996). *Plant functional types and climatic changes: Introduction*. *Journal of Vegetation Science*, Vol. 7, 306-308.
 - Woodward F. I., and M. R. Lomas (2004). *Vegetation dynamics: simulating responses to climatic change*. *Biological Reviews*, 79, 643-670.
 - Xue Y. and J. Shukla (1993). *The influence of land surface properties on Sahel climate. Part I: Desertification*. *Journal of Climate*, 6, 2232-2245.
 - Xue Y. and J. Shukla (1996). *The influence of land surface properties on Sahel climate. Part II: Afforestation*. *Journal of Climate*, 9, 3260-3275.
 - Zaehle S. (2005). *Process-based simulation of the European terrestrial biosphere. An evaluation of present-day and future terrestrial carbon balance estimates and their uncertainty*. PhD Dissertation.
 - Zaehle S., S. Sitch, B. Smith, and F. Hatterman (2005). *Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics*. *Global Biogeochemical Cycles*, 19, GB3020.
 - Zaitchik B. F., A. K. Macalady, L. R. Bonneau, and R. B. Smith (2006). *Europe's 2003 heat wave: a satellite view of impacts and land-atmosphere feedbacks*. *International Journal of Climatology*, 26, 743-769.
 - Zeng N., A. Mariotti, and P. Wetzel (2005). *Terrestrial mechanisms of interannual CO₂ variability*. *Global Biogeochemical Cycles*, 19, GB1016.
 - Zobler L. (1986). *A word soil file for global climate modeling*. NASA Technical Memorandum, 87802, 32.

ACKNOWLEDGEMENTS

I would like to thank Prof. Vincenzo Malvestuto, who has introduced me to the subject of conjugate gradient method and he helped me to develop the inversion algorithms; without his helps I'm still writing the algorithms. I would also thank Dr. Jens Kattge for the Metropolis code and Christian Beer for some useful discussions and suggestions about LPJ.

I wish to thank Prof. Riccardo Valentini for advising my thesis and providing support for my PhD study for the past three years, and Dario Papale for his continuous encouragement and support. A special thank to Dr. Vincenzo Artale and Paolo Ruti for their active role in my education; they allowed me to take an active role in interdisciplinary collaboration.

I am also grateful to Emanuele Lombardi for the constant support he gave me to run the models. Finally, I would like to thank my parents and all my friends.