Analysis of leaf area index in the ECMWF land surface model and impact on latent heat and carbon fluxes: Application to West Africa

L. Jarlan,[1,2] G. Balsamo,[1] S. Lafont,[1,3] A. Beljaars,[1] J. C. Calvet,[4] and E. Mougin[5]

[1]   A new version of the land surface model of the European Centre for Medium-Range Weather Forecasts (Carbon-TESSEL, or CTESSEL) includes a vegetation growth model. This study describes a leaf area index (LAI) data assimilation system (LDAS) based on CTESSEL and satellite LAI for operational Net Ecosystem Exchange (NEE) predictions. The LDAS is evaluated over West Africa. A preliminary experiment shows a significant impact of the LAI on the CTESSEL NEE. The LAI is compared to two satellite products: the predicted annual cycle is delayed over the Sahel and savannah, and the LAI values differ from the satellite products. Preliminary to their use in the LDAS, the LAI products are rescaled to the CTESSEL predictions. The LDAS simulations are confronted to measurements of biomass and LAI for a site in Mali. The LAI analysis is shown to improve the predicted biomass and the annual cycles of the water (latent heat flux, or LE) and carbon (NEE) fluxes. Afterward, the LDAS is run over West Africa with the Moderate-Resolution Imaging Spectroradiometer products (2001–2005). The analysis of LAI shows a limited impact on LE, but it impacts strongly on NEE. Finally, the CTESSEL NEE are compared to two other models' outputs (simple biosphere (SIB) and Carnegie-Ames-Stanford (CASA)). The order of magnitude of the three data sets agrees well, and the shift in annual cycle of CTESSEL is reduced by the LDAS. It is concluded that a LAI data assimilation system is essential for NEE prediction at seasonal and interannual timescales, while a LAI satellite-based climatology may be sufficient for accurate LE predictions.

1.   Introduction

[2]   The land surface conditions are of primary importance for climate and weather prediction. Both root zone soil moisture and vegetation conditions play a vital role in the partitioning of water and energy budgets at the soil-vegetation-atmosphere interface through evaporation processes of the uppermost surface soil layer and plant transpiration [*Shukla and Mintz*, 1982]. This partitioning, in turn, partly controls the thermodynamic and soil moisture content of the lower troposphere, and thus, climate. The quality of short-term to seasonal weather predictions has been shown to strongly depend on a good initialization of the soil moisture [*Beljaars and Viterbo*, 1999]. *Zeng et al.* [1999] and *Philippon and Fontaine* [2001] among others found theoretical evidence of the influence of vegetation on precipitation at interseasonal to interdecadal timescale in the Sahel where the surface-atmosphere feedbacks are known to be large. Apart from partitioning surface energy fluxes, vegetation also governs the natural CO_2 exchanges modulating concentration of atmospheric CO_2 at timescales ranging from minutes to months [*Tucker et al.*, 1986]. Within this context, the land surface models (LSMs) aiming to define the boundary conditions for general circulation models have been recently improved to include an interactive vegetation dynamics [*Arora*, 2002]. Stated differently, the vegetation physiological parameters of interest for the simulation of hydrological processes (leaf area index (LAI) and stomata resistance) are not prescribed and kept constant anymore but evolve interactively with environmental conditions thanks to biophysically based models of vegetation growth and photosynthesis. This approach, although tending toward a more realistic representation of the land surface processes and their interaction, increases the number of uncertain parameters of LSMs and thus, the uncertainties of their simulations. Reducing these uncertainties by combining, optimally, model prediction with observations is the objective of data assimilation (DA).

[3]   Land surface DA has developed supped on the advances introduced by the Numerical Weather Prediction community, a major application being the operational pro-

---

[1]European Centre for Medium-Range Weather Forecasts, Reading, UK.
[2]Also at Centre d'Etudes Spatiales de la Biosphère, Toulouse, France.
[3]Also at Forest Research Agency, Farnham, UK.
[4]Météo-France, Centre National de Recherches Météorologiques/ GMME-MC2, Toulouse, France.
[5]Centre d'Etudes Spatiales de la Biosphère, Toulouse, France.

duction of weather forecast [*Bennett*, 2002]. Application to land surface has received considerable attention within the past few years but most of the studies focused on the improvement of the soil moisture initialization regarding the role of the soil water stock on the quality of the weather forecasts [*Walker and Houser*, 2001]. *Drush* [2007], for instance, shows that the analysis of the predicted root zone from surface soil moisture products derived from satellite observations influences the local weather parameters in the European Center for Medium-Range Weather Forecasts (ECMWF) system. Previous attempts to assimilate satellite-derived vegetation characteristics mainly concerned historical reconstruction. *Knorr and Lakshmi* [2001] shows a better vegetation cover prediction through the identification of uncertain parameters of the BETHY model thanks to the data assimilation of the fraction of photosynthetically active radiation (fAPAR) derived from the AVHRR sensor. *Rayner et al.* [2005] improved this system to reconstruct two decades of terrestrial carbon fluxes thanks to the assimilation of fAPAR and atmospheric CO_2 concentrations. Only little attention has been paid to the improvement of the vegetation characteristics prediction within an operational context apart from the work of *Gu et al.* [2006] who developed an optimal interpolation scheme in the Canadian weather forecast system to provide with LAI at the global scale from the Moderate-Resolution Imaging Spectroradiometer (MODIS) observations. Nevertheless, this last work is based on pure data and does not include a dynamical vegetation model able to simulate carbon fluxes.

[4] Estimation of vegetation characteristics (biomass, LAI) was one of the first quantitative application of remote sensing through empirical relationships established between vegetation indices such as the Normalized Difference Vegetation Index (NDVI) acquired by the Advanced Very High Resolution Radiometer (AVHRR) sensor and ground measurements [e.g., *Tucker et al.*, 1985]. Although strong potentialities were highlighted, these empirical relationships depend on the local conditions of the region where they were developed (soil background, vegetation structure). Furthermore, these relationships do not take into account the effects of water and nutrients limitations on vegetation functioning [*Prince*, 1991; *Lo Seen et al.*, 1995]. The use of radiative transfer models, simulating the physics of the interaction between the electromagnetic waves and the surface components, associated to inversion methods should be theoretically better suited to retrieve a quantitative information on vegetation (LAI, aboveground biomass) from NDVI observations [e.g., *Myneni et al.*, 1997; *Kimes et al.*, 2000]. This approach has been chosen by the MODIS team to extract LAI from measured reflectances [*Tian et al.*, 2000].

[5] The objective of this study is twofold: it aims to (1) demonstrate the interest of LAI data assimilation in order to simulate carbon fluxes and (2) present a LAI data assimilation system for real time prediction of carbon fluxes. The system is based on the LSM of ECMWF named TESSEL [*Van den Hurk et al.*, 2000], recently modified to include an interactive vegetation (hereinafter called CTESSEL for Carbon-TESSEL) [*Lafont et al.*, 2007] constrained by satellite-derived LAI products. The data, the model and the data assimilation method are described in the second section. The third section demonstrates the impact of LAI

on the simulated carbon fluxes and a comparison between the CTESSEL open loop (without DA) and the satellite-derived LAI products is carried out. The results of the LAI data assimilation system over an experimental site located in Mali and at the scale of West Africa are described and discussed in the fourth section. Finally, conclusions and perspectives are drawn.

2.   Materials and Methods
2.1.   Carbon-TESSEL (CTESSEL) Model

[6] The TESSEL model is the land surface model of the ECMWF developed by *Van den Hurk et al.* [2000]. It is designed to describe the exchanges of heat and water between the low-level atmosphere, the vegetation and the soil within atmospheric models. Each grid box can be composed of up to 8 tiles: bare soil, high vegetation, low vegetation, high vegetation with snow beneath, snow on low vegetation, interception layer, sea ice, open water.

[7] TESSEL has been coupled with the A-gs model [*Jacobs*, 1994] that described the control of the latent heat fluxes by the plant and described in a coupled manner the water and carbon cycle. The photosynthesis module A-gs calculates the net CO_2 assimilation from which the biomass and the LAI can be diagnosed. The coupling has been performed by *Lafont et al.* [2007] and *Voogt et al.* [2006] based on the land surface model ISBA-A-gs [*Calvet et al.*, 1998].

[8] The diagnostic of LAI is based on biomass evolution due to photosynthetic activity. The growth module simulates growth and mortality of the vegetation. The growth of active biomass B (in g/m^2) is based on the accumulated net CO_2 assimilation over the previous day $A_{n,\text{day}}$ (see Appendix A for details), and the LAI is obtained from the biomass following

$$\text{LAI} = \frac{B}{\alpha_B}$$

$$\alpha_B = \frac{1}{\text{SLA}} = \frac{1}{eN_L + f}$$

SLA is the specific leaf area in m^2/g of active biomass; e and f are called plasticity parameters and are dependent on vegetation type and N_L is the nitrogen contents of leaves, also dependent on vegetation type. The coupled TESSEL-A-gs is named hereinafter CTESSEL for Carbon-TESSEL. The main equations of the model are described in Appendix A.

2.2.   Data Assimilation Approach: A Simplified Two-Dimensional Variational

[9] Variational methods adjust the model simulations to match the observations available within the assimilation window (at observations times) by minimizing a cost function J, with respect to a background information $\mathbf{x}^b$. The general form of J is given by

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^b)^T \frac{\mathbf{B}^{-1}}{2}(\mathbf{x} - \mathbf{x}^b) + (\mathbf{y} - H(\mathbf{x}))^T \frac{\mathbf{R}^{-1}}{2}(\mathbf{y} - H(\mathbf{x}))$$
$$= J_b(\mathbf{x}) + J_0(\mathbf{x}). \tag{1}$$

The cost function has two terms: the background term $J_b(\mathbf{x})$ which measures the distance between the state vector x and the a priori state $\mathbf{x}^b$ (values of the analyzed variable at the beginning of the assimilation window) weighted by the background error matrix $\mathbf{B}$, and the observation term $J_0(\mathbf{x})$ which accounts for the distance between the vector of observations during the assimilation window, y, and the simulations weighted by the observation error matrix $\mathbf{R}$. The projection of the state vector in the observation space is done through the observation operator $H(\ )$. The minimum of the cost function is given by the classical formulation of the best linear unbiased estimate (with the hypothesis that errors follow a normal distribution and that the linearity hypothesis is fulfilled):

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y} - H(\mathbf{x}^b)), \qquad (2)$$

where $\mathbf{K}$ is called the gain and is calculated as follows:

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T\left[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}\right]^{-1}. \qquad (3)$$

The simplified two-dimensional variational (2DVAR) is an assimilation method developed by *Balsamo et al.* [2004]. The main advantage of the method is to avoid the development of adjoint and linear tangent models to minimize J. It assumes a quasilinear problem close to the background state $\mathbf{x}^b$ by approximating $H(\ )$ by a one-side finite difference. An additional forecast run with perturbed initial condition is required for each state variable to be analyzed, in order to estimate the linearized observation operator $\mathbf{H}$. If we call this two run RUN1 and RUN2, it can be written as

$$\mathbf{H} = \frac{H(x^b_{\text{RUN1}}) - H(x^b_{\text{RUN2}})}{x^b_{\text{RUN1}} - x^b_{\text{RUN2}}}.$$

[10] As such, the simplified 2DVAR is close to an Extended Kalman Filter from which the propagation of the model error matrix is neglected considering a constant $\mathbf{B}$ matrix.

[11] The method was initially designed to analyze the root zone soil moisture using 2-m air temperature and humidity observations. It has been adapted by *Muñoz Sabater et al.* [2007] to analyze root zone soil moisture from surface soil moisture observations and to the analysis of both aboveground biomass and root zone soil moisture by *Muñoz Sabater et al.* [2008]. The simplified 2DVAR has been also applied on the SMOSREX data set (Toulouse, France) using point-scale ground measurements. The method is applied here to the analysis of LAI from satellite-derived products over West Africa. The state vectors $\mathbf{x}$ and $\mathbf{y}$ are respectively the predicted and of the observed LAI at the grid point scale. Analyzing the LAI in CTESSEL is equivalent to the analysis aboveground active biomass (as the ratio between the active aboveground biomass and LAI is constant). The 2DVAR is applied grid point by grid point.

[12] Therefore, $\mathbf{B}$ is diagonal. Finally, the increments $\mathbf{K}(\mathbf{y}-H(\mathbf{x}^b))$ are distributed to the low and high vegetation LAI thanks to their cover fraction F_{low} and F_{high}, respectively.
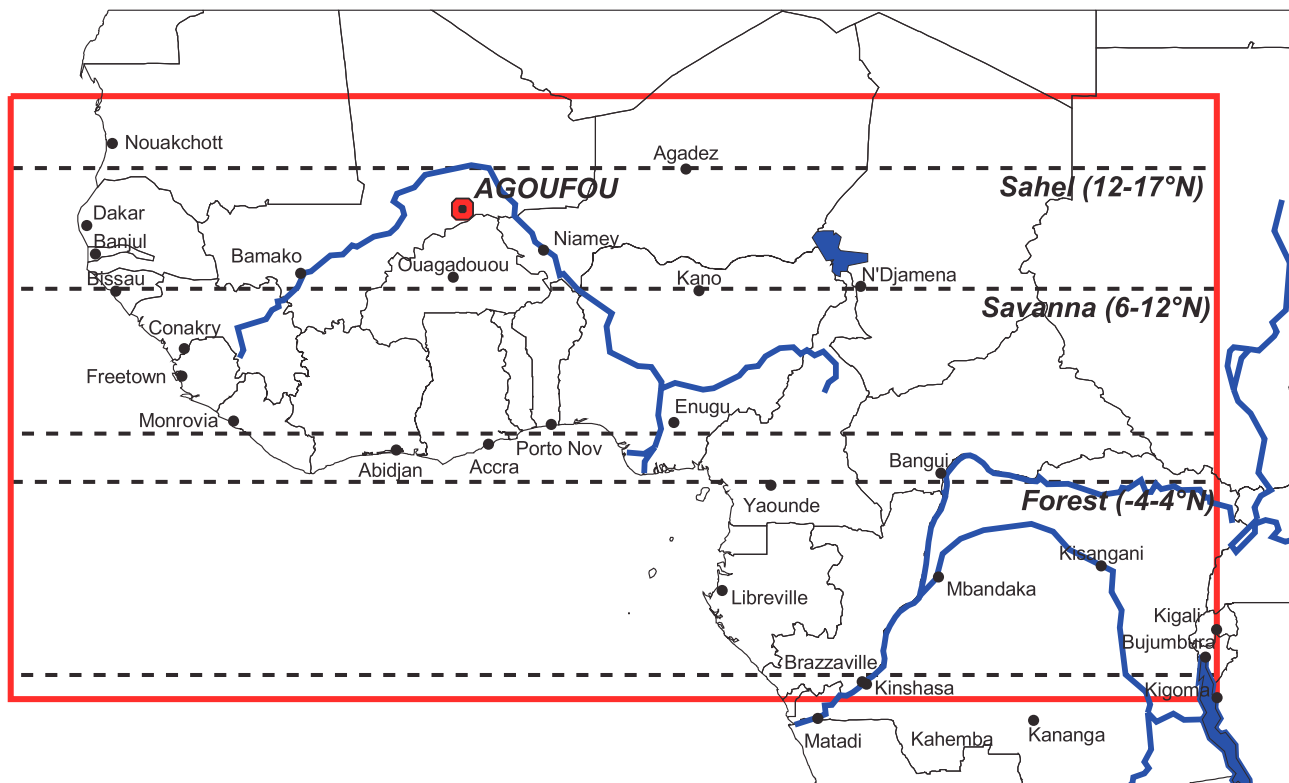
2.3. West Africa Region

[13] West Africa gets most of its annual rainfall during the boreal summer months from June to September. This rainy season is associated with the seasonal reversal of the winds in the lowest level of the atmosphere which is called the monsoon. In winter, the wind blows from the cool continent to the warm ocean. Following the sun apparent movement in the course of the year, the continent warms faster than the ocean. This thermal contrast drives the surface pressure contrast between the ocean (high pressure) and the continent (low pressure) and the set up of the monsoon circulation. The resulting precipitation distribution is characterized by a strong gradient of rainfall amount from north to south and the duration of the wet season range from 1 to 2 months to the border between the Sahara and the Sahel to regularly distributed rainfall events along the year to the south. The annual amount is highly erratic with a higher interannual variability to the north of the window. This climate is locally modulated by the topography (Cameroon mountains, the Adrars to the north). The rainfall gradient (annual amount and length of the wet season) from north to south is associated with a gradient of vegetation (types and fraction cover). Many detailed ecological classification exists for the definition of the bioclimatic zone of West Africa. In this study, the West African is simply subdivided into four bioclimatic zones classified by latitude. These correspond roughly to annual rainfall (PP in mm/a) [*Le Houérou*, 1989]: PP < 150 for the Sahara Desert to the north; 150 < PP < 600 for the African Sahel, which constitutes the transition between the Sahara and the Sudanian savannah (600 < PP < 1500), and PP > 1500 for the Guinean forest to the south. The Sahel is mostly covered by dry savannahs composed of open annual herbaceous vegetation and of a sparse trees and shrubs layer. The Sudanian region is covered by savannah and dry woodlands. Rain forests are concentrated to the south of the window over the Guinean region.

[14] Our study window extends from −5°N to 20°N and from −20°E to 30°W (Figure 1). In addition, three subwindows located in the Sahelian, in the Sudanian and in the Guinean bioclimatic zones, have been defined for sake of clarity of the results presentation.

2.4. Field Data

[15] In order to evaluate our system against ground observations, the Agoufou (15.3°N, 1.3°W) study site has been selected. It is located within the African Monsoon Multidisciplinary Analysis (AMMA) mesoscale site (14.5°–17.5°N, 1°–2°W) in the Gourma region in Mali (Figure 1). The Gourma region is located entirely within the Sahel bioclimatic zone and extends to the South of the Niger River between Timbuctu and Gao down to the border with Burkina-Faso. This is mainly a pastoral region enclosed by the annual average 500 and 150 mm isohyets. The rain distribution is strictly mono-modal with rainfall starting in June and ending in September with a maximum in August. The rainy season is then followed by a long dry season characterized by the absence of green vegetation apart from some scattered trees and shrubs. Rangeland vegetation is composed of an herbaceous layer and a sparse woody plant population. Annual herbs germinate after the first rains, in June or July, and unless the plants wilt before maturity owing to a lack of rainfall, the senescence coin-

Figure 1. The West African windows (black line). The three subregions used for the evaluation of the satellite-derived data and the Carbon-TESSEL (CTESSEL) model are represented by dashed lines: Sahel $(12°-17°N)$, savannah $(6°-12°N)$, and forest $(-4°-4°N)$. The AMMA supersite (near Agoufou village) located in the Gourma region of Mali is superimposed.

cides approximately with the end of the rainy season. The Agoufou site $(1 \times 1 \text{ km}^2)$ is a typical Sahelian landscape characterized by gently undulating sand dunes. The total tree and shrub cover is about 4.5%, whereas the grass cover may vary from 0 to about 60% depending on soil moisture availability. The soil is coarse grained or sandy (>90%).

[16] For the 2005 wet season, the annual rainfall total reached 410 mm which can be considered as a relatively wet year (the long-term average being 370 mm). Ground measurements of the vegetation consist in an estimate of the time variation of LAI and aboveground biomass from trees and grasses using hemispherical photographs [*Weiss et al.*, 2004] and destructive sampling, respectively. For the grass layer, a 1-km transect has been defined in the east-west direction where measurements are performed every 10 m, resulting in 100 pictures. Concerning aboveground herbaceous biomass, 12 samples are used. The large quantity of data is sufficient to capture the spatial variability of the grass layer. The computed mean LAI is assumed to be representative at the 1 km^2 scale. The estimated resulting precision in terms of LAI is 0.23 m^2 m^{-2} (at 1 standard deviation). Further details on the experimental protocol are detailed by *Hiernaux* [1984]. In 2005, the growth of the grass layer started early in June and reached a maximum LAI of 1.8 m^2/m^2 and 2100 kg of dry matter/hectare by the end of August. In contrast, the LAI of trees estimated from pictures taken of isolated individual stands remains at values lower than 0.2 throughout the year.

2.5. Satellite-Derived Leaf Area Index (LAI)

[17] LAI defines an important structural property of a plant canopy as the one sided leaf area per unit ground area. Two LAI products are evaluated and compared to CTESSEL within this study. The first data set, derived from the observations of the VEGETATION instrument, is named CYCLOPES (satellite products for change detection and carbon cycle assessment at the regional and global scales) [*Weiss et al.*, 2007]. The second one is the MODIS LAI [*Tian et al.*, 2000].

[18] The CYCLOPES products are derived from data acquired by the VEGETATION sensor. The VEGETATION instrument on board SPOT4 (launched on April 1998) followed by VEGETATION2 on board SPOT5 (since February 2003) provide with a measure of land surface reflectance in the visible and near-infrared domain continuously. The sensor offers a daily global coverage with a spatial resolution of 1 km^2. Reflectance measurements are performed within four spectral bands. Further details on the instrument characteristics can be found in the work of *Duchemin et al.* [2002]. The LAI products are processed by the GEOLAND Core Service Products for the years 1998 to 2003. The version 3 of the product is used in this demonstrational system. Version 3 algorithm used recalibrated and atmospheric corrected reflectances. The LAI is derived from fCover using the semiempirical approach of *Roujean and Lacaze* [2002]. Further details can be found in the algorithm description [*Lacaze*, 2004] (available

at http://postel.mediasfrance.org/IMG/pdf/CYCL_ATBD-DirectionalNormalisation_I2.0.pdf) and in the work of *Baret et al.* [2008].

[19] The MODIS instrument is operating on both the Terra and Aqua spacecraft. It has a viewing swath width of 2330 km and views the entire surface of the Earth every 1–2 days. Its detectors measure 36 spectral bands between 0.405 and 14.385 μm, and it acquires data at three spatial resolutions: 250 m, 500 m, and 1000 m. The MOD15 LAI is a 1-km global data product updated once every 8 days derived from the MODIS sensor on board TERRA. The MODIS LAI Level 4 algorithms were developed jointly by personnel at Boston University and the University of Montana SCF and NASA GSFC. The algorithm consists of a main procedure based on the inversion of a 3-D radiative transfer model thanks to a look-up table. This algorithm exploits the spectral information content of MODIS surface reflectances at up to seven spectral bands. Should this main algorithm fail, a backup algorithm is triggered to estimate LAI and FPAR empirical relationships from vegetation indices. The algorithm uses a land cover classification that is compatible with the radiative transfer model used in their derivation. The LAI products, in the collection 4 version, are available from 2001 to present.

2.6. Global Net Ecosystem Exchange Data Sets: SIB and CASA

[20] Two global NEE data sets are used for the evaluation of the CTESSEL predictions. The CASA fluxes [*Randerson et al.*, 1997] are based on the Monteith approach: the primary production is the product of the fraction of absorbed radiation (estimated from AVHRR NDVI), the incoming radiation, and an efficiency coefficient. In this simulation the meteorological drivers are based on a climatology. This CASA data set is currently used as boundary condition by the atmospheric transport model at ECMWF and is expected to be replaced by the CTESSEL fluxes. The SiB 3.0 fluxes [*Denning et al.*, 1996] have been computed with forcing from NCEP2 meteorology data. SiB calculates surface fluxes of sensible and latent heat, radiation, moisture, CO_2, and momentum for vegetated land points. The phenological properties (LAI, vegetation cover) of the vegetation are derived from the GIMMS NDVI data sets [*Tucker et al.*, 2005].

3. Implementation of the LAI Data Assimilation System

3.1. Climate, Soils, and Vegetation Data

[21] The climate data are provided by the AMMA Land Surface Model Intercomparison Project (see http://www.cnrm.meteo.fr/amma-moana/amma_surf/almip/index.html). The data consists in standard meteorological variables provided at a 0.5 degree resolution at a 3-hour time step over a West Africa window from 2001 to 2005. This corresponds to a 101 × 51 grid which uses a cylindrical equidistant projection. This first forcing data set, named hereinafter "FORCING1," is based purely on numerical weather prediction (NWP) forecast model output diagnostics (from ECMWF). Additionally, a second forcing data sets is available for years 2004 and 2005 (hereinafter referred to as "FORCING2"). FORCING2 is based on the merging of

FORCING1 and remote sensing-based products. The downwelling longwave and shortwave radiative fluxes for 2004 are from the Ocean and Sea Ice Satellite Application Facility and the precipitation is from Estimation of Precipitation by Satellite (EPSAT) product. In 2005, the downwelling radiative fluxes are from LAND Satellite Application Facility. Figure 2 displays the Hovmöller diagram of cumulative 8-day rainfall for FORCING1 (Figure 2a) and FORCING2 (Figure 2b) over land. Both forcings are consistent with our knowledge of the West African rainfall regime. The north-south gradient of annual amount and length of the wet season are well depicted together with the bimodal regime around the equator [*Leroux*, 1972]. The use of the MSG precipitation products for FORCING2 increases the rainfall amounts over the northern Sahel (above 12°N) during summer. The FORCING2 may be considered as improved with regards to the pure ECMWF forecast as *Andersson et al.* [2007] show a dry bias of the ECMWF rainfall forecast over Sahel.

[22] At the scale of the West Africa, the ECOCLIMAP database is used to fill the CTESSEL input parameters. The ECOCLIMAP data set have been designed to provide a complete set of high-resolution surface parameters for land surface models. It is fully described by *Masson et al.* [2003]. The data set is based on a combination of three products: climate maps, land cover map, and 1 year of AVHRR NDVI data. A classification process per continent is used to assign a homogeneous ecosystem type (214 in total) to each 1-km pixel. All the pixels belonging to a given type have a common land cover. Within this study, the 1-km (214 ecosystem types) ecosystem map is aggregated to the resolution of the meteorological forcing (0.5°) and to a limited number of vegetation types. ECOCLIMAP has been adapted to obtain a seven-type vegetation map: three high-vegetation types (DECIDUOUS, CONIFEROUS, EVERGREEN) and four low-vegetation types (C3 GRASS, C4 GRASS, C3 CROPS, and C4 CROPS). Finally, each grid point can be composed of two kinds of vegetation type (low vegetation and high vegetation) with associated fraction cover F_{low} and F_{high}, respectively.

[23] The ECOCLIMAP data set provides a number of parameters for SVAT models, including albedo, minimum stomatal resistance, roughness lengths and fraction of soil cover. The photosynthesis parameters of the A-gs module (g_c, N_a, e, f, etc.; see Appendix A) are taken from *Gibelin et al.* [2006] who describe the implementation at the global scale of ISBA-A-gs and validate the global simulation from satellite observations. In addition, ECOCLIMAP provides with a climatology of monthly Leaf Area Index that can be used to forced the LAI in CTESSEL by shunting the growth module. The ECOCLIMAP LAI is computed by linearly mapping the minimum NDVI and the maximum NDVI of the AVHRR data during the season 1992–1993 to the minimum LAI and the maximum LAI per ecosystem gathered from the literature. Further details are given by *Masson et al.* [2003].

3.2. 2DVAR Setup
#### 3.2.1. Background and Observation Errors

[24] The two different available forcing data sets are used to evaluate the diagonal terms of the background error covariance matrix **B**. CTESSEL with the diagnosed LAI
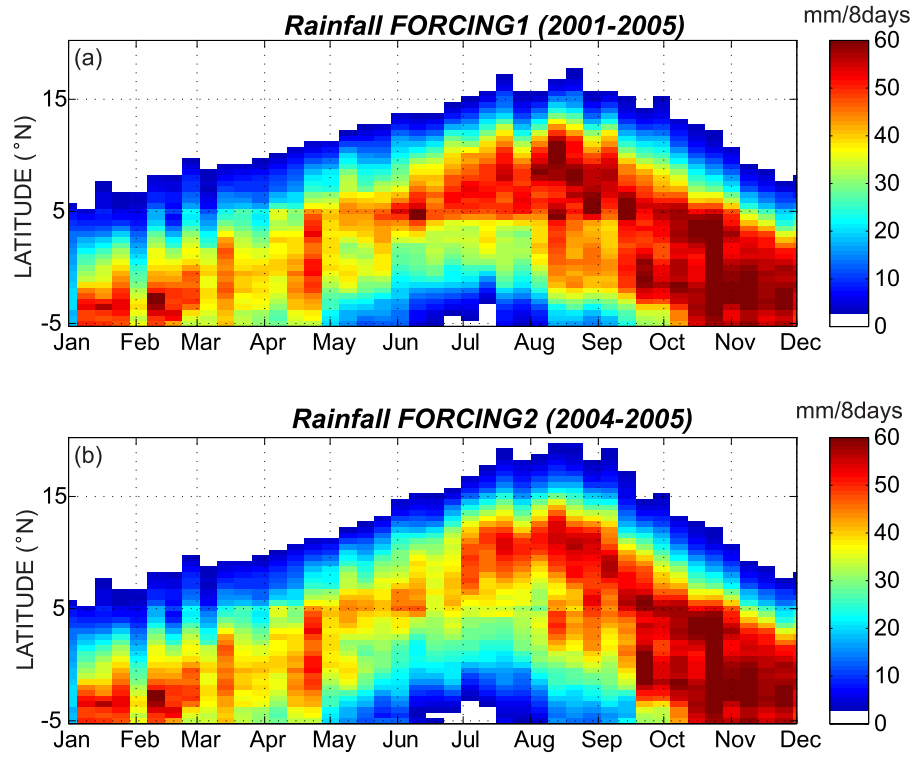
Figure 2. Eight-day rainfall Hovmöller diagram for (a) FORCING1, averaged over 2001–2005, and (b) FORCING2, averaged over 2004–2005, over land.

is run with FORCING1 and FORCING2 for the years 2004 and 2005. These two experiments are named EXP1_CT and EXP2_CT, respectively. The average annual cycle of the LAI differences between the two experiments are plotted at Figure 3 as Hovmöller diagram.

[25] The analysis of Figure 3 can give us indication of the background error. The differences in terms of precipitation between the two forcings are located between 5° and 15°N (Figure 2). Therefore, we focus on this region where the LAI differences are obviously the larger (Figure 3). The
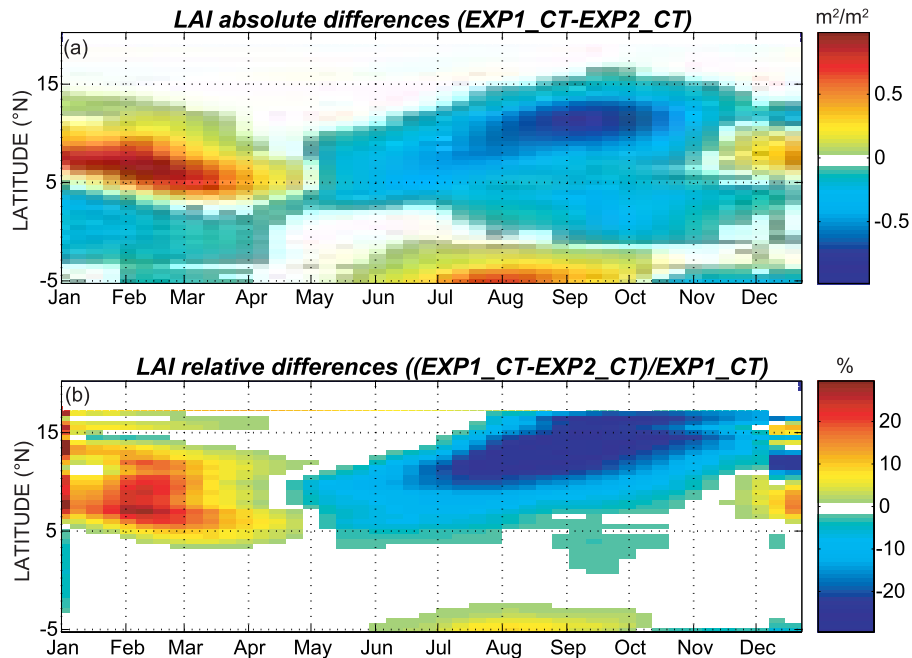
Figure 3. Average annual cycles of the differences between EXP1_CT and EXP2_CT (see text) for leaf area index (LAI) over 2004–2005: (a) absolute differences (m^2/m^2) and (b) relative differences (%).
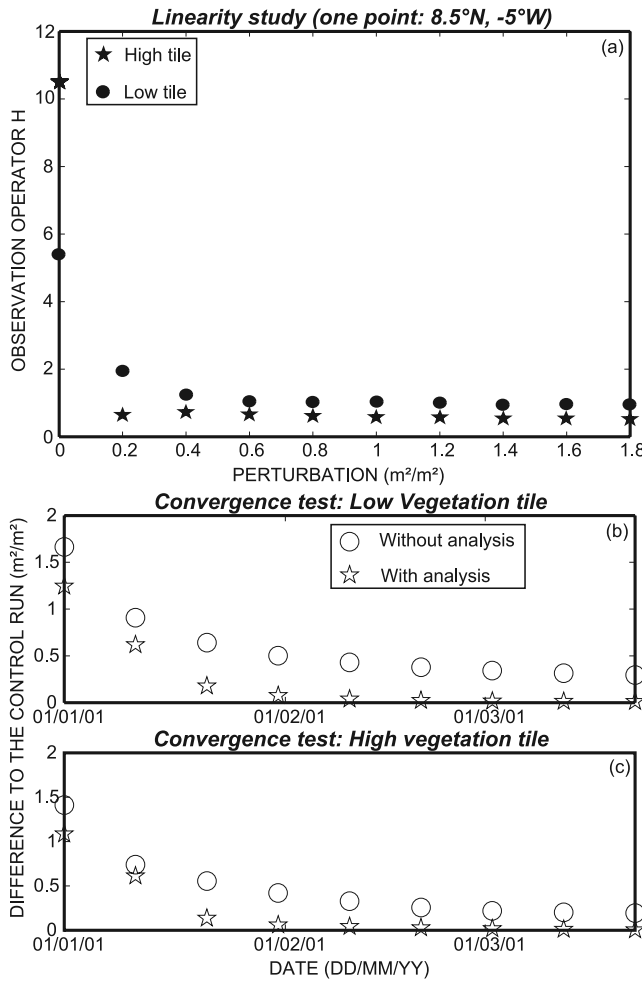
Figure 4. (a) Test of the linear hypothesis for the simplified 2DVAR method on one point. Apart from numerical errors due to low perturbations, the system is very linear for perturbations chosen above 0.4 m²/m²; (b, c) Convergence test (absolute difference between the open loop (circles) and the analysis (stars)) and the control run for the low- and high-vegetation tiles, respectively. The system converges in about 1 month for the analysis, whereas it takes a longer time for the open loop (without LAI assimilation).

lower precipitation of FORCING1 lead to lower simulated EXP1_CT LAI than for the EXP2_CT experiment. In addition, these LAI differences are mainly associated to the growth phase of the vegetation during July, August and September. During the first half of the year, EXP1_CT LAI is higher than EXP2_CT LAI. This may be attributed to the later rainfall events of FORCING1 than FORCING2 that keep quite high LAI until late in the winter. The absolute and relative differences between the two experiments range from −0.9 to +1.2 m²/m² and −29 to +27%, respectively. From these elements, the background error is set semi-empirically to 20% of forecasted LAI with low and high bounds at 0.2 and 1.2 m²/m² over the whole domain of study. Nevertheless, the use of this rather empirical and spatially homogeneous model errors statistics is a first approximation and better estimations, variable in space

and time, should be considered in the future. The analysis ensemble method [*Kucukkaraca and Fisher*, 2006] which is based on an ensemble of forecasts seems a promising technique.

[26] The satellite-derived LAI uncertainty proceeds from different sources (atmospheric and directional effects badly corrected on the reflectance used in the inversion process, inversion errors, spatial representativity, etc.). Therefore, the errors of satellite-derived LAI are difficult to estimate and ecosystem-dependent. Within this study, we compute the average time series of MODIS LAI over the "forest" window (Figure 1). Assuming this is a stable target, we apply the ergodicity principle and fix the observations errors to two standard deviation of this time series (i.e., about 1 m²/m²). Finally, errors are assumed to be uncorrelated between two 8- to 10-day satellite acquisitions.

3.2.2. Linearity Test and Choice of the Perturbation

[27] The linear approximation used to numerically compute the observation operator has to be evaluated. In order to test the linear hypothesis, an ensemble of 10 perturbations from 0.01 to 1.8 m²/m² is computed and the observation operator is calculated for all the grid points over land.

[28] Apart from numerical errors for small perturbations, the results show that the sensitivity of the calculated **H** does not depend on the size of the perturbation above a threshold (around 0.4 m²/m²). The observation operator is plotted against the size of the perturbation for one grid point at Figure 4a but similar conclusion are drawn for other points. The perturbation of LAI is taken equal to 0.4 m²/m².

3.2.3. Evaluation of the Analysis Convergence

[29] The convergence of the assimilation system is evaluated using twin experiments. A reference run initialized with the January LAI from ECOCLIMAP is used to produce a set of LAI simulated "observations" that mimics the satellite temporal repetitivity (10 days). The initial conditions are then modified such as $LAI_{low} = LAI_{high} = 3.0$ over the window. A free run (without LAI assimilation) and a run with the analysis of LAI using the artificial observations are computed. Figure 4b and c shows the RMSE between these two runs and the reference run for the high- and the low-vegetation tile, respectively. The results show a good convergence of the data assimilation approach after four cycles (40 days) for both high- and low-vegetation types. The free run also converges toward the reference run. As the forcing used is the same, both runs tend toward the same equilibrium. Nevertheless, the convergence is much slower for the free run and it still does not fit the control run after 3 months of integration.

3.2.4. Length of the Assimilation Window

[30] At ECMWF, the land data assimilation system will include soil moisture and LAI analysis. Regarding the temporal scale of variability of soil moisture, the shorter the window, the better. *Calvet* [2000] and *Muñoz Sabater et al.* [2007] have shown the good performance of a 10-day assimilation windows for soil moisture data assimilation. Concerning LAI, the variability is much smoother and several tests have been done by extending the window length up to 50 days (not shown). The results in terms of simulated NEE and LE are not very sensitive to the window length. Finally, the length of the assimilation window is the same as the temporal repetitivity of the satellite-derived
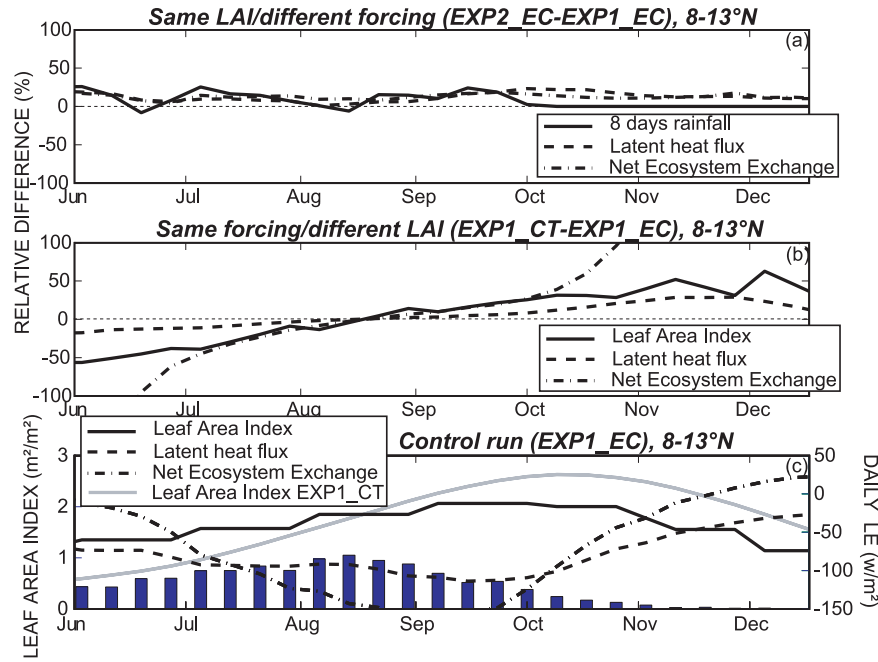
Figure 5. (a) Average cycles of the relative differences between EXP2_EC and EXP1_EC (CTESSEL with a climatology of LAI forced by FORCING1 and FORCING2, respectively; see text) for rainfall, Net Ecosystem Exchange (NEE), and LE from 8° to 13°N. (b) Average cycles of the relative differences between EXP1_CT and EXP1_EC (CTESSEL forced by FORCING1 with a diagnostic and a climatology of LAI, respectively; see text) for LAI, NEE, and LE from 8° to 13°N. (c) The CTESSEL prediction for the control run (EXP1_EC): the convention for fluxes is positive upward. LE (W/m^2) is multiplied by -1, and NEE (μmol/m^2/s) is multiplied by 30, to match the scale of LE fluxes on the right axis (Figure 5c). Eight-day rainfall (divided by 50 to be read on the left axis) and EXP1_CT LAI have been superimposed for comparison purposes. The differences are normalized with respect to EXP1_EC.

products (8 days for MODIS and 10 days for CYCLOPES), thus including one observation per window.

4. Preliminary Experiments

[31] In this part, we intend to demonstrate the interest of analyzing the Leaf Area Index in the CTESSEL model by quantifying the impact of variations of LAI on the simulated LE and NEE. Second and as a preliminary step toward data assimilation, the free runs of the CTESSEL model (without DA) are compared to the satellite-derived LAI.
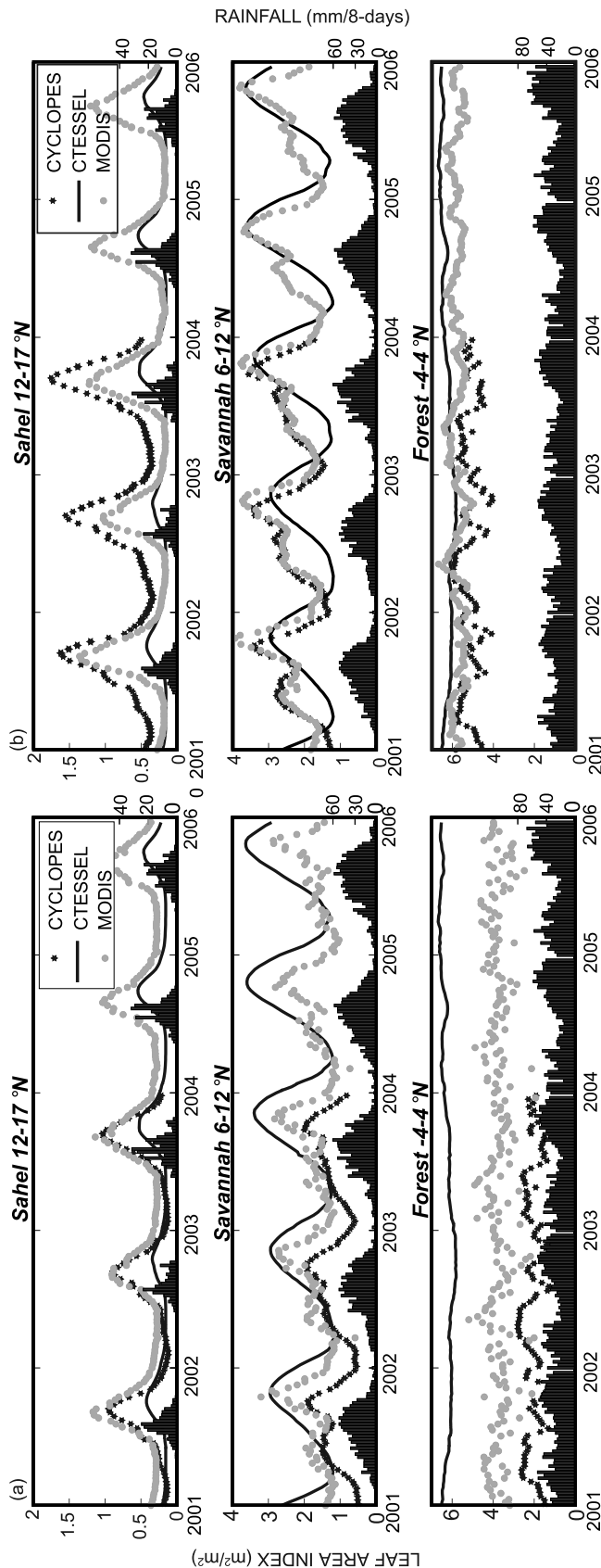
4.1. Impact of the LAI on LE and NEE Predictions

[32] In this part, we take benefit of the two forcings to discriminate the impact of the forcing from the impact of the LAI on the NEE and LE. The FORCING2 is characterized by a higher rainfall annual amount than FORCING1, particularly over the northern part of the window (see Figure 2). We will concentrate on a large region covering southern Sahel and northern savannah (between 8°N and 13°N). The CTESSEL model is run for the overlapping period of the two forcing (years 2004 and 2005) with FORCING1 and FORCING2 and with a diagnosed LAI. These two experiments, named EXP1_CT and EXP2_CT, are similar to those used above to evaluate the background error. In addition, two other runs are computed where the LAI is not diagnose anymore but taken from a climatology (ECOCLIMAP) [*Masson et al.*, 2003] and is equal for the two runs. These two additional experiments are named

EXP1_EC and EXP2_EC, respectively. The average annual cycles over years 2004 and 2005 of the differences between the EXP2_EC and EXP1_EC (same LAI but different forcing) in terms of rainfall, NEE and LE are calculated. Likewise, the differences between the EXP1_CT and EXP1_EC (same forcing but different LAI) in terms of LAI, NEE and LE are calculated. The results are normalized with respect to the so-called control run EXP1_EC. Figure 5a shows the normalized differences (in percent) between EXP2_EC and EXP1_EC averaged from 8°N to 13°N from June to October which corresponds to the wet season (where the differences between the two forcings are the most important). Likewise, Figure 5b shows the normalized differences (in percent) between EXP1_CT and EXP1_EC. Figure 5c shows the EXP1_EC results (averaged from 8°N to 13°N) in terms of LAI, daily NEE, daily LE and 8-day rainfall. All over the paper, the convention for fluxes is positive upward. Nevertheless, for sake of clarity of Figure 5c, LE (W/m^{-2}) has been multiplied by -1 and NEE fluxes (μmol CO$_2$ m^{-2} s^{-1}) by +30 to match the scale of LE fluxes on the right axe. Eight-day rainfall (divided by 50 to be read on the right axes) and EXP1_CT LAI have been superimposed for comparison purposes.

[33] The rainfall difference between FORCING2 and FORCING1 reaches 38% and is about 16% on average between the beginning of July and mid-September (Figure 5a). In terms of fluxes, the higher rainfall of FORCING2 leads obviously to an increase of both NEE

Figure 6. Time series of LAI by latitudinal bands (from top to bottom, Sahel, savannah, and forest). (a) Comparison of satellite-derived LAI (MODIS and CYCLOPES) before preprocessing (see text) with CTESSEL open loop; CTESSEL and MODIS are plotted from 2001 to 2005, and CYCLOPES is plotted from 2001 to 2003. (b) Same as Figure 6a, but after rescaling for both satellite data sets and temporal smoothing for MODIS. Additionally, 8-day rainfall amounts are given by black bars (scale is on the right axis).

and LE (12% and 11% on average over the same period). As previously underlined, this rainfall differences decrease to zero after the end of September. The differences in terms of fluxes remains late in the season but are not really significant as the absolute values of LE and LEE are low (Figure 5c). The difference of behavior between the EXP1_CT (interactive LAI) and EXP1_EC (climatology of LAI) experiments are of particular interest (Figure 5b) regarding the impact of LAI on the fluxes. The LAI cycle of CTESSEL is delayed of about one month with regards to the climatology (Figure 5c). In particular, the peak LAI occurs at the beginning of September for the climatology whereas it occurs one month later for CTESSEL. The difference in terms of LAI between EXP1_CT and EXP1_EC is −27% on average between the beginning of June and mid-August. From mid-August to the end of December, it is inverted and equal to +24% on average. In terms of fluxes, LE and NEE behaves differently to these differences in LAI. Whereas the impact of LAI on LE remains low (−7% and +11% over the two periods identified above), LAI appears to be the main factor governing the annual cycle differences of the NEE fluxes between these two experiments. The relative differences between EXP1_CT and EXP2_CT are −39% from the beginning of July to mid-August and +22% from mid-August to mid-October and follows closely the LAI differences curve. As a conclusion, both LE and NEE predictions are obviously influenced by rainfall. The LAI impacts in a different way. Whereas its influence on LE remains limited, this study demonstrates that the phenology of LAI strongly governs the time where the surface shift from sources to sink during the growth phase of the vegetation, and inversely, from sink to sources during the senescence phase.

4.2. CTESSEL Against the Satellite-Derived LAI and Bias Correction

[34] Before using the satellite LAI products within the assimilation scheme, a comparison between CTESSEL and the two satellite-derived data sets is performed. In order to reduce the established bias between CTESSEL and the satellite products, a bias correction process described in Appendix B is carried out. The time series of LAI for the CTESSEL, MODIS and CYCLOPES LAI are plotted in Figures 6a and 6b averaged for the three latitudinal bands before and after bias correction, respectively. Precipitations are superimposed for comparison purposes. Only grid points where the satellite products are available are taken into account. FORCING1 is preferred to FORCING2 for the CTESSEL runs as it covers the period of availability of the satellite-derived products (2001–2005 for MODIS and 2001–2003 for CYCLOPES). In addition, the FORCING1

corresponds to the pure outputs of the ECMWF forecasting system. This allows to be placed in real condition for the future LAI data assimilation system to be implemented at ECMWF.

[35] Over Sahel and savannah (Figure 6a), the three data sets exhibit a strong seasonality associated to the displacement of the ITCZ to the north during the summer months. The north-south gradient of the length of the growing season is also well marked with longer season over savannah than Sahel. Over forest, the signals are more stable all over the year even if the satellite-derived LAI are more scattered than the CTESSEL one. This can be attributed to a bad correction of atmospheric, directional effects and/or cloud detection on the reflectances used to retrieve LAI. In contrast, the annual cycles are differently depicted. In particular, the peaks of LAI simulated by CTESSEL are lagged of about one month over Sahel with regards to the two satellite products. Over savannah, they are in phase with MODIS whereas CYCLOPES shows a LAI peak one month earlier. This delay between CTESSEL and the satellite products has also been observed by *Gibelin et al.* [2006] who evaluate ISBA-a-gs (from which the photosynthesis and growth modules are taken) at the global scale. The decrease of LAI after the end of the rainy season is quicker for CYCLOPES than for MODIS. The absolute values of LAI are also very different. CYCLOPES shows the lower values over savannah and forests. CTESSEL LAI is clearly the higher below 12°N and MODIS presents intermediate values between the two others. Over forest, the overall stable LAI between 6 and 7 m^2/m^2 simulated by CTESSEL is much higher than the satellite products. In particular, the CYCLOPES LAI over forest are far too low (<3 m^2/m^2). Furthermore, the MODIS data sets appear strongly scattered with regards to CYCLOPES, in particular over savannah and forest. The standard deviations over forest are equal to 0.55 m^2/m^2 and 0.41 m^2/m^2 for MODIS and CYCLOPES, respectively. Over Sahel, the vegetation is mainly composed of annual herbs and the LAI during the dry season should be close to zero [*Le Houérou*, 1989]. CYCLOPES and CTESSEL exhibits these expected low LAI whereas MODIS shows significantly higher values. In contrast, the satellite-derived LAI are strongly overestimating the CTESSEL simulations during the wet season over Sahel. This can be attributed to the already described dry bias of FORCING1 over this region. Over savannah, a decrease of LAI centered on August is striking for MODIS and, to a lesser extent for CYCLOPES. This could be attributed to the well known little dry season affecting these areas during August and September but it has been associated to the increase of the cloud cover at this period of the year by *Moulin et al.* [1997]. The absence of a significant decrease of rainfall during this period tends to corroborate this assumption (Figure 6).

[36] The statistical characteristics of LAI from satellite and CTESSEL differ substantially. Furthermore, the simplified 2DVAR, similarly to others DA methods derived from the Best Linear Unbiased Estimates, aims at correcting for Gaussian errors with a mean equal to zero and not for systematic errors. At this time, a global map of LAI toward the model and the observations could be corrected does not exist. Prior to assimilation of the satellite retrievals into the CTESSEL model, satellite model biases are thus removed by scaling the satellite retrievals into the land model prediction through matching of their respective histograms. This approach has been developed by *Reichle et al.* [2007] for soil moisture observations. This matching procedure aims at reducing the bias between the model and the observations without making any assumption of which part of the bias comes from the model or from the data. In addition, a smoothing method is applied to the MODIS observations. The bias correction and the smoothing processes are described in Appendix B. The time series of satellite products after smoothing and bias correction are more in agreement with CTESSEL (Figure 6b). Over forest in particular, the overall stable LAI of about 6 m^2/m^2 matches the model prediction. The satellite products are still affected by a low seasonal signal that is more to be attributed to atmospheric perturbations than to real vegetation phenology. The scattering of MODIS LAI has been strongly minimized thanks to the smoothing. Over savannah, the two satellite data sets are in good agreement all together and matches the CTESSEL LAI values in terms of the minimum and maximum but the annual cycle, as already highlighted above, strongly differs. Over Sahel, the amplitude of the annual signal is much higher for the satellite products than for CTESSEL. In addition, because of the rescaling, the dry season LAI value for CYCLOPES is higher than the simulated CTESSEL LAI whereas for MODIS, it matches better. Finally, the phenology the satellite products (date of the beginning of growth and date of peak LAI) are not modified substantially by the rescaling (see Appendix B).

5. Data Assimilation Results and Discussion

5.1. Point-Scale Evaluation

[37] The 2DVAR is evaluated on the Agoufou site during the growing season 2005. Ground LAI measurements, extracted every 10 days to mimics satellite repetitivity, are assimilated into CTESSEL in order to analyze aboveground biomass. The tree cover fraction being low, trees are not considered in this study and CTESSEL is run considering only C4 grasses. Furthermore, the implementation of the data assimilation system with C3 grasses, C3 crops or C4 crops does not modify the conclusions of the study. Figure 7 displays the evaluation of the 2DVAR over the Agoufou site: LAI, aboveground biomass, midday latent heat fluxes and midday NEE averaged over 10-day periods are displayed in Figures 7a–7d, respectively. The convention for fluxes is positive upward.

[38] This site, located in the Sahelian zone sensu stricto as defined by *Le Houérou* [1989] is characterized by a short rainy season of about three months. The vegetation growth, composed mainly of annual herbs, is very rapid and is followed by a strong senescence after the last rainfall events. This is well depicted by the LAI and aboveground biomass measurements (Figures 7a and 7b). The open loop simulations of CTESSEL do not match this short growing period. This is in accordance with the findings of the comparison between CTESSEL LAI and the satellite LAI products (over Sahel the simulated LAI peak is delayed with regards to the satellite products). Over the Agoufou site, the CTESSEL peak is also delayed by about 25 days together with the senescence phase that extents further after the last
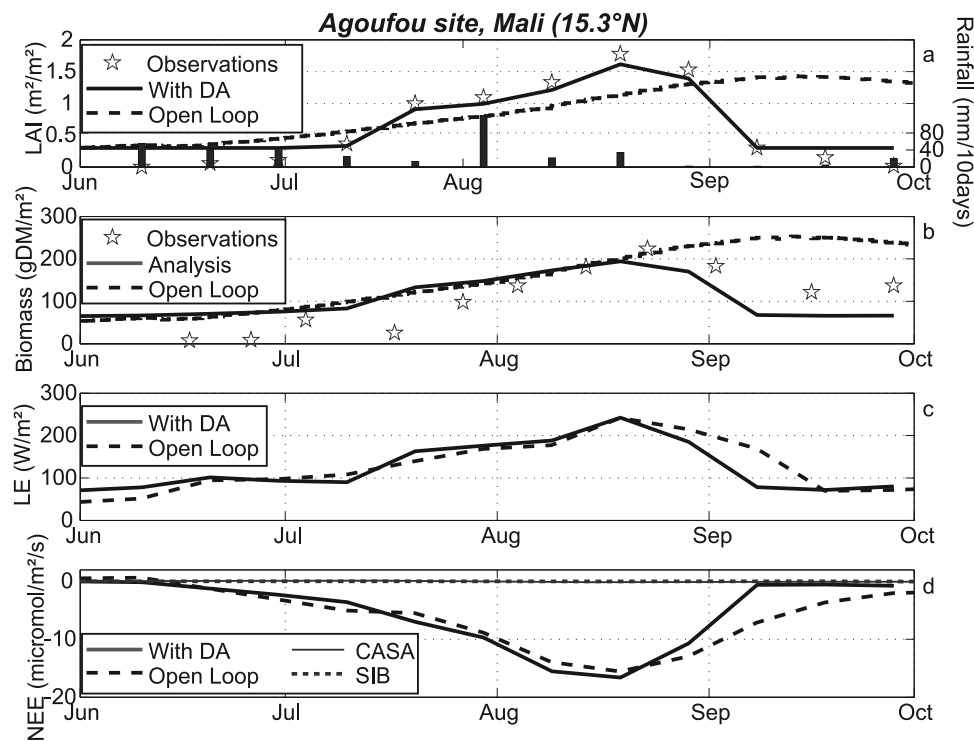
Figure 7. A 2DVAR analysis of aboveground biomass from LAI observations over the Agoufou site (Mali) during the year 2005: (a) LAI (solid line, after data assimilation; dashed line, open loop; stars, observations); (b) aboveground biomass (solid line, analysis; dashed line, open loop; stars, observations); (c) 10-day average latent heat flux at midday (solid line, after data assimilation; dashed line, open loop; stars, observations); and (d) 10-day average NEE at midday (solid line, after data assimilation; dashed line, open loop; stars, observations). SIB and CASA NEE of the corresponding grid points are superimposed. Convention for fluxes is positive upward. The open loop CTESSEL simulations (dotted line) and 10-day precipitation (bar) are superimposed for comparison purposes.

rainfall events during the dry season. In contrast, the value at LAI peak is quite well simulated by CTESSEL (1.45 m²/m² and 1.80 m²/m² for CTESSEL and ground measurements, respectively). This differs from the comparison with the satellite products but, over Agoufou, the meteorological forcing is measured on site whereas over West Africa, the FORCING1 based on the ECMWF forecast is used. The ECMWF operational forecasting system is known to underestimate the Sahelian precipitation. The measured annual rainfall is 410 mm/a whereas the corresponding grid point value for FORCING1 is 47 mm/a (against 286 mm/a for FORCING2). Concerning biomass, the 2DVAR data assimilation system also corrects the delay of CTESSEL open loop but the decrease phase appears slightly in advance.

[39] The analysis of aboveground biomass and LAI has a low impact on water and NEE during the growing phase (see Figures 7c and 7d). In particular, the peak value of latent heat flux (LE) is very close with and without data assimilation (DA) (242 W/m² and 241 W/m², respectively). *Sivakumar* [1990] underlines that the primary factor governing evaporation variability over Sahel are the temporal patterns of rainfall. The impact of the 2DVAR on the NEE peak is also limited (-15.45 and -16.59 μmol CO_2/m²/s with and without DA, respectively). During the senescence phase, differences are more striking. The CTESSEL open loop maintains a high evaporation rate and a high NEE even after the last rainfall events. This can be attributed to the

higher LAI and consequently, the higher water extraction from depth soil by the roots. To that respect, the CTESSEL with the LAI analysis has a behavior better in agreement with what we know of the LE dynamic over Sahel. *Wallace and Holwill* [1997] among others show a return to very low evaporation rate just 2 days after the last rainfall events thanks to LE measurements performed at the end of the rainy season over a Sahelian site similar to the Agoufou site. For information, SIB and CASA NEE has been superimposed in Figure 7d. Both show almost no variability with fluxes around 0 all around the growing season. These two data sets forced by AVHRR NDVI for vegetation characteristics may suffer from the limitation of this sensor in case of low vegetation cover. CTESSEL NEE may represent a strong improvement over the Sahelian region with regards to these two data sets.

[40] For lack of available LE and CO_2 fluxes on the site, the order of magnitude of CTESSEL simulations is just roughly verified by comparison to measurements already performed by other teams over similar ecosystems. In terms of LE, the daily (24-hour mean) value (averaged over a 10-day period) reaches 90 W/m² (not shown) that corresponds to around 3 mm H_2O/d. This is in good agreement with value usually given for evapotranspiration over Sahel during the core of the rainy season such as the measurements performed during the Hapex-Sahel Experiment [*Kabat et al.*, 1997]. Concerning NEE, it shows a clear

Table 1. Peak Values of LAI, LE, and NEE for CTESSEL Open Loop and CTESSEL Analysis for Different Values of Precipitation[a]

	LAI (m^2/m^2)	LE (W/m^2)	NEE ($\mu mol\ CO_2/m^2/s$)
	PP		
Open loop	1.43	241.6	−15.45
2DVAR	1.62	242.3	−16.59
	PP/2		
Open loop	1.09	152.2	−7.95
2DVAR	1.62	180.1	−9.09
	PP/4		
Open loop	0.7	81.2	−4.31
2DVAR	1.62	117.2	−7.94

[a]Abbreviations are as follows: 2DVAR, two-dimensional variational; CTESSEL, Carbon-TESSEL; LAI, leaf area index; LE, latent heat fluxes; NEE, Net Ecosystem Exchange; PP, measured precipitation; PP/2, measured precipitation divided by 2; PP/4, measured precipitation divided by 4.

diurnal patterns (not shown) with the highest uptake rates around noon and a permanent slight release to the atmosphere at night. This is corroborated by measurements of *Hanan et al.* [1998] performed over a natural savannah. They give noon peak values reaching, on certain days, −15 $\mu mol\ CO_2/m^2/s$ (for LAI peak equal to 0.9 m^2/m^2). *Falk et al.* [2007] measured on a site of natural savannah located in Burkina noon values of −22 $\mu mol\ CO_2/m^2/s$ on average over July and August (for LAI peak equal to 3.2 m^2/m^2; C. Brüemmer, personal communication, 2008). CTESSEL with LAI analysis gives −16.6 $\mu mol\ CO_2/m^2/s$ (average over a 10-day period and for a simulated LAI peak of 1.45 m^2/m^2; compare Figure 7).

[41] Concerning daily values, *Hanan et al.* [1998] measurements reaches −3 $\mu mol\ CO_2\ m^2/s$ whereas CTESSEL with LAI analysis is a significantly higher value of −4.54 $\mu mol\ CO_2\ m^2/s$ (not shown) as it is an average over a 10-day period. This may be attributed to the soil respiration which is parameterized from the surface temperature without taking into account soil moisture. *Hanan et al.* [1998] values of nighttime respiratory CO_2 fluxes are between 2 and 5 $\mu mol\ CO_2\ m^2/s$ (average value around 3 $\mu mol\ CO_2\ m^2/s$). In contrast, CTESSEL (with or without LAI analysis) nighttime CO_2 fluxes are around 2 $\mu mol\ CO_2\ m^2/s$ which is quite low. In particular, *Veenendaal et al.* [2004] underline the occurrence of marked spikes of CO_2 release because of rainfall events over a semiarid savannah in southern Africa. This behavior is also observed over the Agoufou site after the analysis of the first measurements of NEE performed at the beginning of the growing season 2006 (V. Le Dantec, CESBIO, personal communication, 2008). This suggests the need for a better parameterization of the soil respiration into CTESSEL in the near future. Finally, the return to very low value of the simulated NEE after the last rainfall events has already been observed by *Verhoef et al.* [1996].

[42] Two additional experiments have been carried out to investigate the effect of negatively biased forcing precipitation (such as FORCING1) on the simplified 2DVAR system. Table 1 displays the peak values of LAI, LE and NEE for the open loop simulations and the 2DVAR system. Three cases are considered: the initial measured precipitation, precipitation divide by 2 and precipitation divided by

4. In contrast to the results with the initial forcing precipitation, LAI, LE and NEE peak values strongly differs between 2DVAR and open loop simulations by introducing a dry bias on the precipitation forcing. In particular, if the precipitation is divided by 4, the 2DVAR system is able to maintain an LAI close to the one simulated with the initial forcing. Furthermore, the LE and NEE are divided by 2 whereas they are divided by 3 and about 4 for the open loop, respectively. As a conclusion, the impact of the analysis of LAI on LE and NEE is low if the water availability is sufficient for vegetation growth (apart from the delayed decrease of the fluxes after the last rainfall). When the precipitation is negatively biased, the analysis of LAI allows maintaining a significantly higher evaporation rate and NEE than with the open loop. This could be of particular interest for the simulation of the African Monsoon by the NWP model of ECMWF.

5.2. Application to the West Africa Window

[43] The 2DVAR algorithm has been evaluated using twin experiments and over a site where ground measurements were available. The objective of this part is to apply this algorithm to the real MODIS LAI products (after rescaling and smoothing). The impact of the LAI data assimilation on the simulated latent heat and NEE is evaluated. Results are compared to the TESSEL model (without the interactive vegetation part) in terms of LE for information.

5.2.1. LAI Time Series

[44] Figure 8 displays the LAI time series for the three latitudinal bands described above. Thanks to the 2DVAR system, the low values of LAI over Sahel are corrected with peak values ranging from 1.0 to 1.5 m^2/m^2 instead of values lower than 0.5 m^2/m^2 for CTESSEL open loop. Over savannah, the analysis follows very closely the MODIS products apart from the spring months where analysis are slightly above the satellite LAI. Indeed, during the forward integration of the model after the analysis, LAI is decreasing (not shown). This is attributed to the simulated environmental conditions (and in particular the root zone soil moisture) that are not sufficient to maintain such a high LAI. Around the LAI peak, analysis and open loop are more in agreement. Finally, the senescence phase is closely following the last rainfall events of the year for the analysis whereas it is too slow for the open loop (this corroborates the remark made over the Agoufou site above). Another striking feature is also the artificial break in the growing phase of the LAI for the MODIS product during the spring months. It is certainly an artifact of the satellite products due to bad atmospheric corrections [*Moulin et al.*, 1997] as previously discussed. Nevertheless, it has been minimized thanks to the smoothing filter with the upper envelope. Over forest, the LAI are quite stable for the two LAI apart from noise attributed to atmospheric perturbations of the MODIS products.

5.2.2. Quality of the Analysis

[45] For diagnosing the properties of data assimilation systems, we use three quantities: (1) the differences ($\mathbf{y}-H(\mathbf{x}^b)$ in equation (2) are called innovations, (2) the residuals, computed by replacing the background value $\mathbf{x}^b$ by the analyzed one $\mathbf{x}^a$ in the differences above. It represents the fit of the observations to the analysis, and finally (3) the analysis increment, reconstructed a posteriori by subtracting
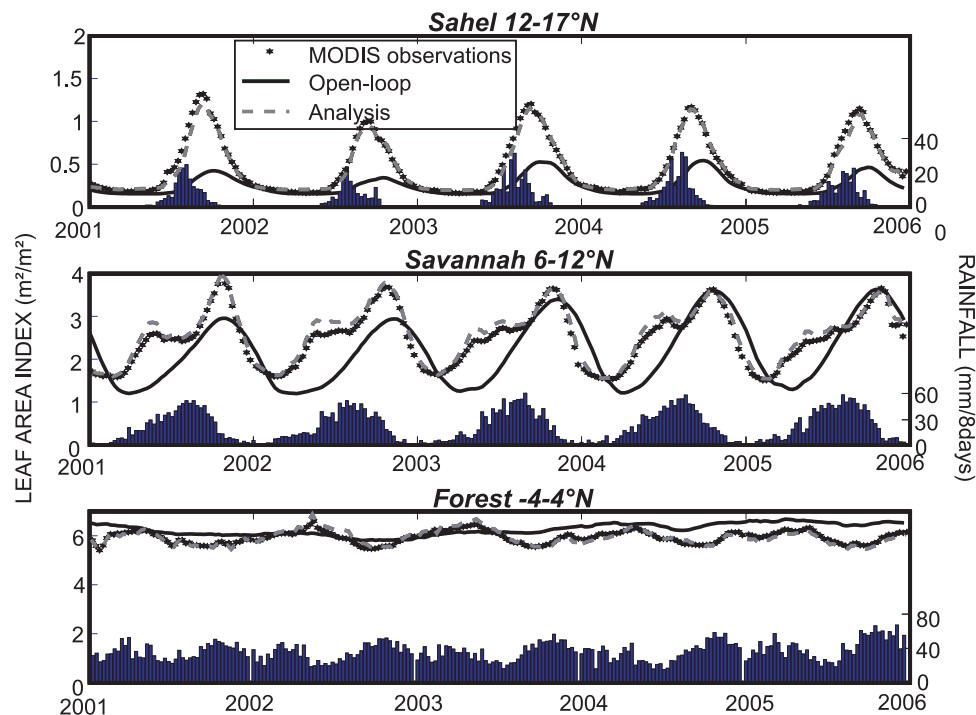
Figure 8. Time series of LAI by latitudinal bands (from top to bottom, Sahel, savannah, and forest): CTESSEL analysis and MODIS. Additionally, 8-day rainfall amounts are given by black bars.

the background from the analysis. Within this study, the analyzed and the observed one are the LAI. By lack of ground "truth," these three quantities can reveal a lot of the performance of the assimilation system. Figure 9 displays the Hovmöller diagram (average over 2001–2005) of the innovations (Figure 9a), the residuals (Figure 9b), and the analysis increments (Figure 9c). Values below 0.2 m^2/m^2 have been masked for a better clarity of the figure.

[46] First, the fit of the observations to the background and analysis can be conveniently examined by Figures 9a and 9b. As expected, one can see that the differences are smaller for the analysis departures (residuals) than for the background departures (innovations). Stated differently, the LAI analysis draw the CTESSEL simulation to the MODIS products thanks to the simplified 2DVAR. Information has been extracted from the observations. Furthermore, most of the time, the innovations and residuals are low (below < 0.3 m^2/m^2). Below 5°N, corresponding to equatorial forest areas, this is attributed to the quite constant LAI both for CTESSEL and for the observations together with the favorable climatic conditions (in terms of rainfall) to maintain such high LAIs (see Figure 8). Over savannah and Sahel, the residuals and the innovations are also negligible during winter because of the low and constant LAI values corresponding to the dying vegetation of these dry months. The innovations and the residuals are significantly different from zero (tested with a t test) during spring over savannah and during summer over Sahel. For these two cases, a strong negative bias of CTESSEL has already been identified even after the rescaling process. This means that the analysis is above the observations at the beginning of the assimilation window but, at the time of the observations (around the middle of the assimilation window), it has gone below the observations. Indeed, the rainfall and consequently,

soil moisture, conditions are not sufficient to maintain such a high quantity of vegetation and the vegetation is dying during the forward integration of the model.

[47] In a perfect data assimilation system increments should be small and time average close to zero which would illustrate an absence of bias between the model and the observations. The preprocessing of the data has strongly minimized the bias as shown above (Figure 6b), and the analysis increment (not shown). Nevertheless, some bias still appear over savannah and Sahel. This bias is attributed to (1) the slow growth in the CTESSEL model (e.g., the delayed peak of LAI values in Figure 6b) and to (2) the dry bias of the precipitation used to force the model, in particular over Sahel. Nevertheless, it is interesting to point out that even in a case of a biased forcing, the assimilation system is still available to provide with realistic value of LAI. As a conclusion, the system is performing quite well with residuals lower than innovations and analysis increments reasonably small over most of the study window.

[48] Figure 10 displays the average analysis increments over the period 2001–2005. The map (Figure 10a) exhibits a marked contrast between savannah and Sahel to the north with positive increments and forest on the south where the increments are slightly negative. The histogram (Figure 10b) should be centered on zero in the case of unbiased data but it is clearly shifted to positive values. Nevertheless, this shift remains reasonable (mean = 0.05 m^2/m^2) and it has been strongly reduced thanks to the a priori rescaling of the satellite products.

5.2.3. Effect of the LAI Analysis on Simulated Latent Heat and CO₂ Fluxes

[49] The objective of introducing an interactive LAI and a data assimilation system in the model of ECMWF is twofold: (1) vegetation, and LAI in particular, affects, the
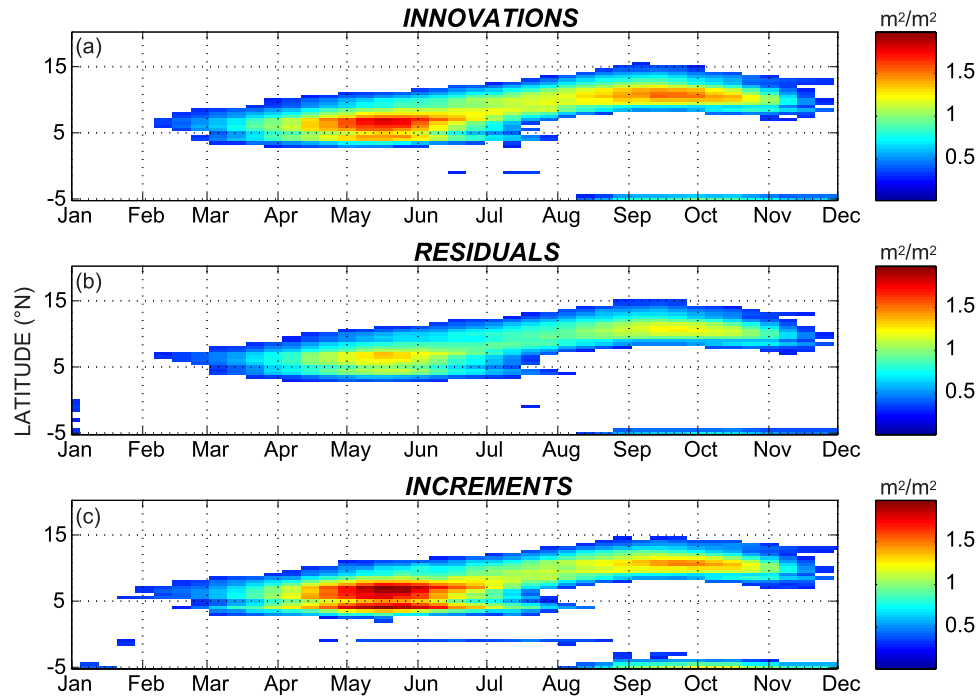
Figure 9. Hovmöller diagram of (a) the innovations, (b) the residuals, and (c) the analysis increments (see text; average annual cycle over the period 2001–2005).

surface latent heat flux (*van den Hurk et al.* [2003], among others, demonstrate the impact of LAI on the annual land surface evaporation cycle) that is of primary importance for NWP; (2) to provide with NEE prediction based on a biophysically based model (a photosynthesis and a vegetation growth modules).that will be propagated into the atmosphere thanks to the atmospheric dynamical model.

[50] Figure 11 displays the annual cycle of the differences (in percent) between CTESSEL+LAI analysis and CTESSEL open loop (divided by CTESSEL open loop simulations) in terms of LAI, LE and NEE averaged over 2001–2005. The LAI differences between the two runs are obviously consis-

tent with what has been underlined at Figure 6b. The main LAI differences (on the order of 50% more for CTESSEL analysis) are located over Sahel in summer and over savannah at spring. On average, this 50% LAI leads to a similar increase in NEE except at the beginning of spring between 7° and 10°N where the increase is higher than 80%. CTESSEL LAI reach their lower values of the year (about 1 m^2/m^2, see Figure 6b) whereas the analysis draw the simulations to the MODIS products values, about 1.5 m^2/m^2 higher. The NEE encountered at this time of the year is low (because of a high incoming radiation combined with a high water stress). This explains the high relative differences in NEE. Likewise, the
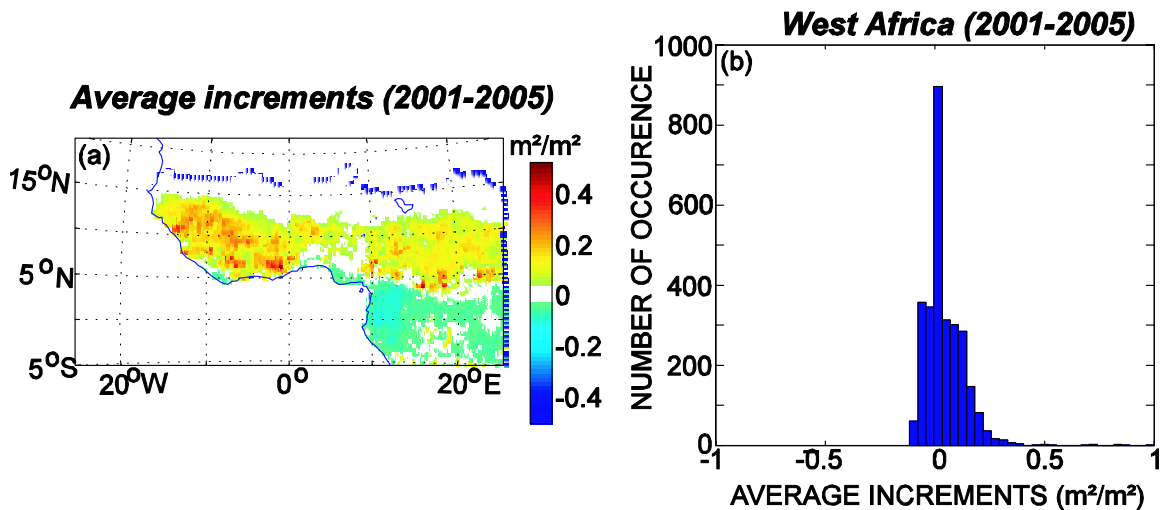
Figure 10. Average analysis increments over 2001–2005: (a) map of West Africa and (b) histogram of values.
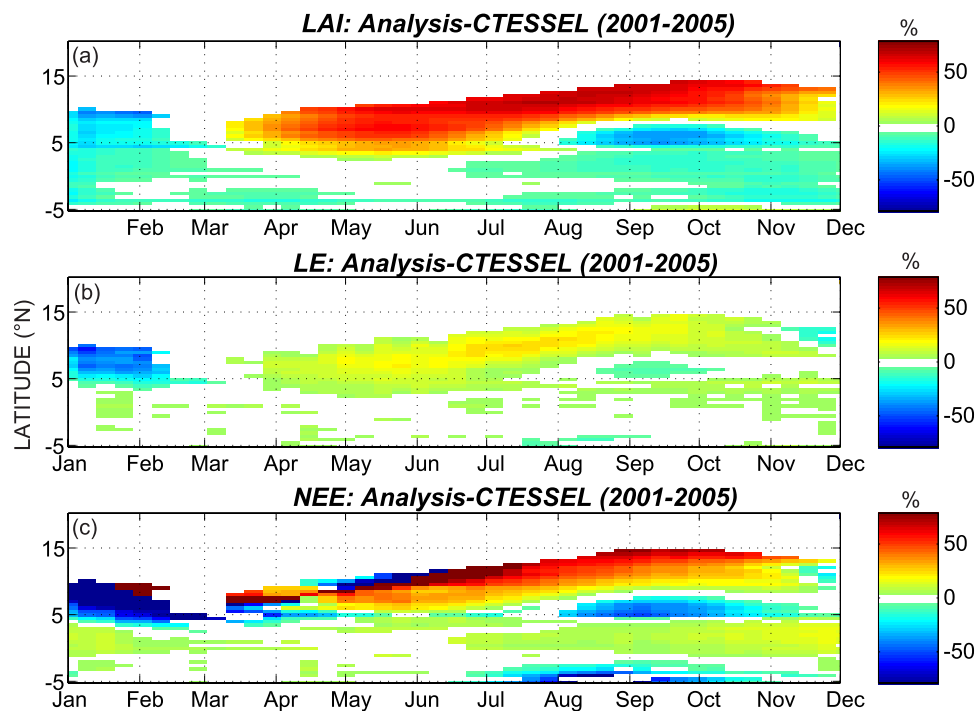
Figure 11. Relative difference in percentage between CTESSEL + LAI analysis and CTESSEL open loop averaged over 2001–2005 with FORCING1 for (a) LAI, (b) LE, and (c) NEE. Relative differences ranging from −3% to +3% are masked for clarity purposes.

−30% LAI differences around January over savannah are also reflected in a significant decrease in NEE (on the order of −80%).

[51] The impact on LE is much more moderate with a maximum impact of about +24% (corresponding to a LAI difference of +70%). In addition, this maximum impact occurs at the beginning of the rainy season when the absolute values of LE are low. Most of the time, the impact of LAI on LE is lower than 15% (with LAI differences of more than +50%). In addition, the difference reach around −40% over a large area extending between 5°N and 10°N at the end of the rainy season (around January). This is attributed to the shifted senescence phase of CTESSEL open loop already observed over the Agoufou site (see Figure 7).

[52] In order to investigate the impact of LAI analysis on LE and NEE interannual variability, the normalized standard deviation (i.e., the standard deviation divided by the average) over 2001–2005 of LAI, LE and NEE are computed and summarized by Hovmöller diagrams in Figure 12 for open loop and CTESSEL + Analysis. The normalized standard deviation is hereinafter improperly called "variability" for simplicity. Hovmöller of rainfall and TESSEL LE variability are also displayed for comparison purposes in Figures 12g and 12h, respectively. For information, the LAI used for TESSEL is a constant LAI all over the year of 3 m^2/m^2 and 5 m^2/m^2 for low and high vegetation, respectively. The variability of 8-day rainfall (Figure 12h) ranges from 22% to 223%. This range of values is much more than for other plotted variables because of the short temporal scale of variability of rainfall. The highest variability is encountered over Sahel. Rainfall, in the Sahel is mostly generated by squall lines, which typically arrive at 3-day intervals throughout the rainy season. Furthermore, storms within these squall lines are convective and the temporal distribution of rainfall is thus highly variable. This high temporal variability combined to low 8-day amount gives high normalized standard deviation (above 100%). Another area of high variability is located to the extreme south of the area during the summer months. During this period of the year, the ITCZ reaches its northern position leading to a significant decrease of the 8-day rainfall amount to the south and consequently, an increase of its variability.

[53] The Hovmöller diagrams of LE for CTESSEL open loop, analysis and TESSEL (Figures 12c, 12d, and 12g, respectively) presents quite similar shape. Furthermore, most of the patterns are in good agreement with the ones of precipitations (Figure 12h). In particular, the northern strip of LE high variability from September to January and from April to June can be related to the similar pattern of the 8-day rainfall. These two temporal periods correspond to the beginning and to the end of the rainy season, respectively. During the core of the rainy season, the rainfall events are more regularly distributed (Figure 12h) and LE displays a lower variability on the three runs (CTESSEL open loop (Figure 12c), CTESSEL analysis (Figure 12d), and TESSEL (Figure 12g)). By contrast with the northern area, the area of high rainfall variability to the south during the summer months (already identified above) is not accompanied by a strong LE variability. The high annual rainfall amount allows for a full filling of the soil reservoir. Consequently, even during the short period of lower rainfall, water availability is sufficient to keep high evaporation rate. The day-to-day variability is also further minimized thanks to the high LAI encountered over these region (around 6 m^2/m^2;
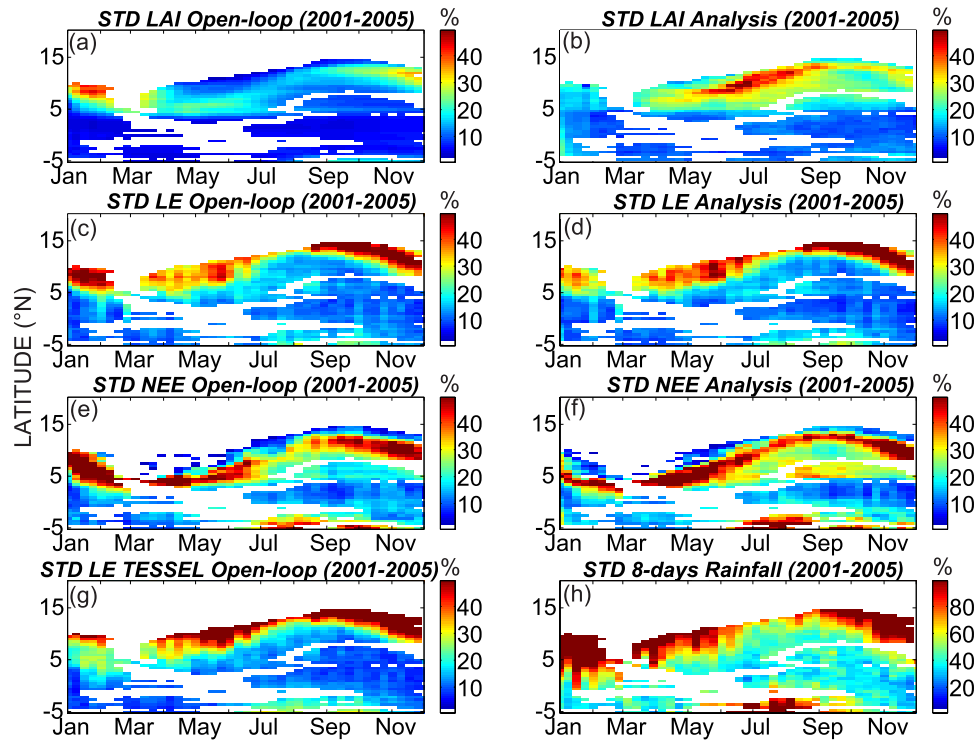
Figure 12. Hovmöller diagrams of the (8-day average) normalized standard deviation (standard deviation divided by the mean) over the 2001–2005 period for LAI, LE, and NEE in percentages: CTESSEL open loop (left) and CTESSEL analysis (right). The normalized standard deviation of TESSEL LE (constant LAI) over the same period and the standard deviation of the 8-day cumulative rainfall are displayed in Figures 12g and 12h, respectively, for comparison purposes. Areas where LAI analysis and abs(LAI analysis - LAI open loop) are below 0.2 m^2/m^2 are masked.

see Figure 8). In addition, the LAI is not fluctuating a lot (see Figures 12a and 12b).

[54] In contrast with LE, the LAI and NEE Hovmöller diagrams exhibits more distinct differences between CTESSEL open loop and CTESSEL analysis. In addition, the Hovmöller patterns are not following so closely the rainfall patterns of variability. The northern strip of rainfall variability does not exhibit a high LAI and NEE variability. Over these areas, LAI and the assimilation of carbon by the vegetation (which is mainly governing the NEE) are low and their year-to-year variability is also below 20%. Just south this strip of low variability, LAI and NEE displays their highest variability. This area of high variability on both LAI and NEE is moving northward with time (corresponding to the northward movement of the ITCZ). The variability of LAI analysis during the core of the rainy season over Sahel and northern savannah (above 10°N) is higher than the one of LAI open loop. The variability of NEE follows quite closely the variability of LAI with a higher variability of NEE analysis than for the NEE open loop from May to December. In contrast with LE, the variability of NEE is strong even during the core of the rainy season over Sahel (August/September). This period of the year corresponds to the LAI peak and potentially the highest year-to-year absolute differences in terms of LAI. Nevertheless, the influence of the rainfall on NEE variability is also important. The area of high rainfall variability to the south during the summer months (already identified above) is accompanied by a high variability of NEE (in contrast with LE). Vegetation may strongly decrease the assimilation of carbon during stress period even if the LAI is high. The assimilated carbon will be used to maintain vegetation tissues alive and will not create new tissues. The resulting NEE will decrease. This could explains the highest variability of NEE over these areas.

[55] As a conclusion, the analysis of the normalized standard deviation of LAI, LE and NEE shows that the interannual variability of LE is more related to rainfall variability than to LAI whereas NEE variability is in good agreement with LAI variability. This has strong consequences for land surface and, in particular, vegetation modeling for NWP and climate: Figures 11 and 12 show that, regarding the limited impact of analyzing LAI on LE, a LAI data assimilation appears of secondary importance for the simulation of LE. Inversely, an interactive LAI reacting to environmental conditions combined with a data assimilation system to avoid model divergence is needed for NEE prediction at the seasonal and interannual timescales.

5.3. Comparison of CTESSEL NEE With SIB and CASA

[56] The CTESSEL NEE with and without LAI data assimilation is compared to two terrestrial surfaces CO_2 flux data set: the CASA and SIB fluxes. The NEE simulated by CTESSEL open loop and analysis averaged over the years 2001 to 2005 are plotted together with SIB and CASA
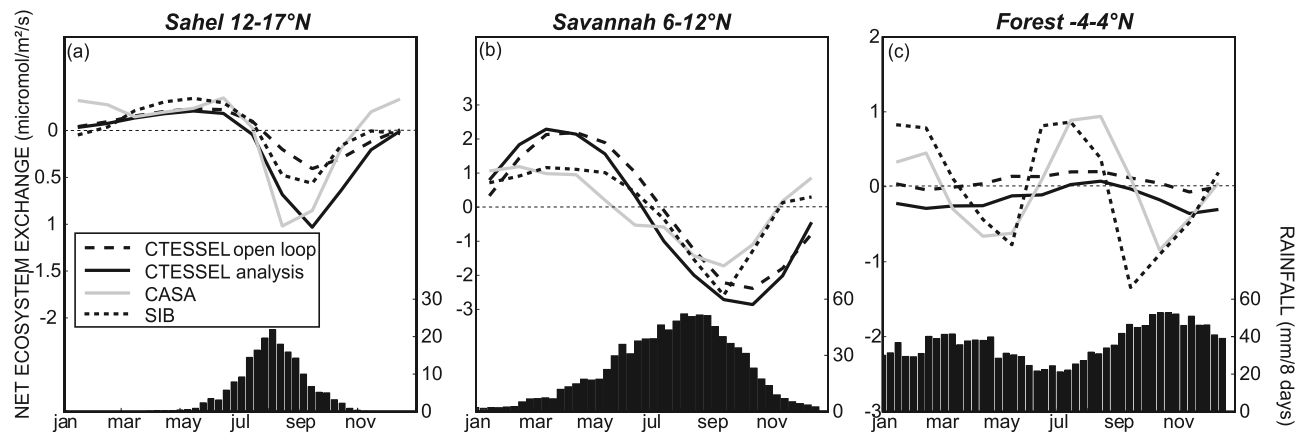
Figure 13. Monthly NEE by latitudinal bands for CTESSEL open loop and analysis (climatology from 2001 to 2005), SIB, and CASA (see text): (a) Sahel, (b) savannah, and (c) forest.

simulations at Figure 13 for three latitudinal bands. CASA is a climatology (simulated with average meteorological and NDVI forcings) and the available SIB fluxes are for year 2002. Table 2 summarizes the NEE average value over the "growing" season (criterion: NEE < 0) and the "dry" season (criterion: NEE > 0) for the four data sets.

[57] Above all, the four NEE data sets are of the same order of magnitude. Over Sahel, CTESSEL with LAI data assimilation leads to higher NEE during the growing season than CTESSEL open loop because the LAI is increased by the analysis (compare Figure 8), and consequently is the assimilation of CO_2 by the plants. During the dry season, CTESSEL open loop and analysis match together as NEE is mainly governed by the soil respiration (parameterized by a function of the surface temperature in model). CASA and SIB respiration take into account soil moisture. This may explain the differences observed with CTESSEL. Notably, at the beginning of the growing season, the first rainfall events are known to cause CO_2 drop into the atmosphere because of soil respiration (V. Le Dantec, personal communication, XXXX). This behavior seems to be caught by CASA and SIB whereas CTESSEL exhibits a smooth increasing shape from December to June to be associated with the regular increase of the incoming solar radiation (not shown). During the growing season, the assimilation of CO_2 by the plants (A_n term in equation (A12)) drives the NEE. Differences in the A_n calculation together with differences in the meteorological forcing used explain quite high discrepancy between the four data sets. Nevertheless, the marked seasonality associated to the African monsoon is well reproduced by the three models. Concerning the wet season NEE (Table 2), CTESSEL open loop and CASA are the lower, SIB shows the higher value whereas CTESSEL analysis is in between. Over the period of study, the annual cycle of CTESSEL open loop and analysis is delayed with regards to the two other data sets. The peak NEE, that occurs in September for CTESSEL, is delayed of about month with regards to SIB. Concerning the beginning of the growing season, the analysis of LAI in CTESSEL helps to correct for this time lag. Over savannah, the four data sets display peaks NEE of the same order of magnitude. CTESSEL (analysis and open loop) has a stronger seasonal signal than the two others with higher respiration rates

during the dry season and higher CO_2 absorption during the growing season. Compared to SIB and CASA, the CTESSEL seasonal cycle is also characterized by a CO_2 assimilation peak delayed of one month as over the Sahel. The decrease of the CO_2 assimilation after the peak occurs systematically later for CTESSEL than for SIB and CASA. These shifted seasonal cycles are in accordance with the delayed LAI cycle already identified above. The LAI data assimilation partly improves this shifted CO_2 fluxes. And in particular, the time when the surface goes from source to sink of carbon at the beginning of the growing season. Over forest, SIB and CASA are slightly shifted in time but they both show a strong biennial signal (with CO_2 absorption during spring and autumn and CO_2 drop during summer and winter) that is not reproduced by the CTESSEL model either with or without LAI analysis.

[58] As a conclusion, CTESSEL exhibits CO_2 fluxes in relatively good agreement with two data sets used as boundary condition for CO_2 transport model. After data assimilation, CTESSEL NEE is in better timing with SIB and CASA than CTESSEL open loop. At ECMWF, CTESSEL is intended to replace the CASA climatology used for the CO_2 surface fluxes. Its ability to simulate

Table 2. Average Monthly NEE Values by Season for CTESSEL Open Loop, CTESSEL With LAI Data Assimilation, SIB, and CASA[a]

	CT OL	CT AN	SIB	CASA
Dry Season, NEE > 0				
Sahel	0.13	0.13	0.21	0.23
Savannah	1.50	1.49	0.72	0.77
Forest	0.18	0.12	0.56	0.41
Growing Season, NEE < 0				
Sahel	−0.25	−0.43	−0.21	−0.69
Savannah	−1.42	−1.83	−1.44	−1.07
Forest	−0.15	−0.25	−0.78	−0.57

[a]See text. The dry season is defined as NEE > 0 and the wet season as NEE < 0. Average monthly NEE values are in $\mu mol/m^2/s$. Abbreviations are as follows: CASA, Carnegie-Ames-Stanford approach; CT AL, CTESSEL with LAI data assimilation; CT OL, CTESSEL open loop; SIB, simple biosphere.

interannual variability of surface carbon fluxes is expected to bring a significant improvement on the actual system.

6. Conclusions and Perspectives

[59] The ECMWF land surface model has been modified to include an interactive vegetation module able to simulate CO_2 Net Ecosystem Exchanges and to diagnose Leaf Area Index (LAI). The model is named CTESSEL for Carbon-TESSEL. A LAI data assimilation system to simulate operationally NEE is presented. It is based on the CTESSEL model constrained by satellite-derived LAI products. The evaluation of the system is performed over West Africa from 2001 to 2005. A preliminary experiment shows a significant impact of the LAI on NEE predictions from CTESSEL open loop over a region covering a large part of Sahel and savannah. Prior to the data assimilation, two satellite-derived LAI data sets are compared to the CTESSEL open loop simulations over West Africa: the CYCLOPES and the MODIS LAI. The satellite products are significantly lower than the CTESSEL predictions for the equatorial forests and the CTESSEL annual cycle is shifted with regards to satellite-derived LAI over Sahel and savannah. The growth is quite low and the senescence is late. The bias between the satellite-derived LAI and the CTESSEL open loop is reduced thanks to a histogram matching method. Finally, the MODIS LAI are chosen as input for the data assimilation system as they are available over the entire period of study and they fit better to the CTESSEL simulations after matching. The data assimilation method is a simplified 2DVAR. It presents the advantage to avoid the development of the adjoint and the tangent linear of the CTESSEL model.

[60] This approach is first evaluated locally on the Sahelian site of the AMMA project located in Mali. With regards to the open loop, the CTESSEL simulations constrained by the analysis of LAI are more in agreement with the ground measurements in terms of LAI and aboveground biomass. In particular, the delayed growth of the open loop is corrected thanks to the data assimilation system. In terms of fluxes, open loop and analysis are close until the LAI peak. Afterward, the open loop maintains a high evapotranspiration rate and NEE thanks to a high LAI. In contrast, the analysis has reduced the LAI to zero: the evapotranspiration and NEE are much lower and close to nil. Additionally, the order of magnitude of NEE and LE (with or without data assimilation) are in good agreement with measurements performed by other teams over similar ecosystems. Nevertheless, the soil respiration parameterization of the CTESSEL model is too rough to capture the dynamics of respiration of the Sahelian soils and must be improved in a near future.

[61] The simplified 2DVAR approach is applied to West Africa in a second part. It is shown that the data assimilation performs well in terms of LAI with analysis closer to the satellite-derived LAI than the forecast. Therefore, the slow growth and the late senescence of the CTESSEL open loop LAI simulations highlighted above are strongly improved thanks to the data assimilation system. In terms of fluxes, even if the forcing, and in particular precipitations plays a major role on LE and NEE, the analysis of LAI impacts much stronger the annual cycle and the interannual variability of NEE than LE. Finally, the NEE predictions before and after analysis of LAI are compared to two other models output: CASA and SIB. The order of magnitude of the three data sets agrees well and the delayed annual cycle of CTESSEL is reduced by the LDAS. It is concluded that a LDAS is essential for NEE predictions at seasonal and interannual timescales while a LAI satellite based climatology may be sufficient for accurate LE predictions.

[62] At ECMWF, the CTESSEL model is intended to provide the surface fluxes of natural CO_2 for the atmospheric transport modeling in the near future and to substitute the current operational land surface scheme once the global verification will be completed. The future operational land data assimilation will include together the analysis of the root zone soil moisture [see *Drush*, 2007] and the analysis of LAI presented in this paper. As such, the system will simulate operationally NEE. In addition, the evaluation of CTESSEL is carrying on within the frame of AMMA Land Surface Model Intercomparison project. Finally, the soil parameterization will be improved by introducing a model of soil respiration in a near future and the simplified 2DVAR for the analysis of LAI will be evaluated at the global scale.

Appendix A: CTESSEL Model

A1. Photosynthesis and Stomatal Resistance Model

[63] The canopy resistance is calculated from photosynthesis, which is the: net CO_2 assimilation (A_n) by the canopy. A_n is calculated as a function of different environmental factors. First, CO_2 assimilation limited by the air CO_2 concentration is determined via a saturation equation:

$$A_m = A_{m,\max}\left(1 - e^{\left(-g_m(C_I - \Gamma)/A_{m,\max}\right)}\right), \quad (A1)$$

where $A_{m,\max}$ is the maximum net CO_2 assimilation, g_m is the mesophyll conductance, C_i is the CO_2 concentration in the leaf and is the CO_2 concentration at which assimilation compensates respiration, called CO_2 compensation concentration. $A_{m,\max}$ depends on temperature via a Q_{10} function. The internal CO_2 concentration C_i, is directly derived from the CO_2 concentration in the air C_s. It is controlled by the air humidity via the specific humidity deficit of the air D_s. If the deficit exceeds the maximum deficit tolerated by the vegetation $D_{\max}$, the plant closes its stomata. $D_{\max}$ is vegetation-dependent. The CO_2 assimilation limited by CO_2 concentration is further limited by radiation by

$$A_n = (A_m + R_d)\left[1 - e^{\frac{-\varepsilon I_a}{(A_m + R_d)}}\right] - R_d, \quad (A2)$$

where I_a is the photosynthetic active radiation (PAR), ε is the initial quantum use efficiency and R_d is the dark respiration. R_d is parameterized simply as a function of A_m.

[64] The stomatal conductance to CO_2, g_{sc}, is estimated using a flux gradient relationship, modified to account for the effect of a specific humidity deficit on stomatal aperture. The first guess g_{sc}^* is given by

$$g_{sc}^* = \frac{A_n - A_{\min}\left(\dfrac{D_s}{D_{\max}}\dfrac{A_n + R_d}{A_m + R_d}\right) + R_d\left(1 - \dfrac{A_n + R_d}{A_m + R_d}\right)}{C_s - C_i}, \quad (A3)$$

where $A_{\min}$ represents the residual photosynthesis rate (at full light intensity) associated with cuticular transfers when the stomata are closed because of a high specific humidity deficit. The diffusion of CO_2 interacts with that of water vapor. The first guess of the stomatal conductance to CO_2 is be corrected for this interaction and calculated interactively to refine the estimation. Finally, the stomatal conductance to water vapor g_s is given by

$$g_s = 1.6 g_{sc} + g_c, \tag{A4}$$

where g_c is the cuticular conductance (a vegetation-dependent parameter).

[65] The soil stress parameterization is based on a meta-analysis of several herbaceous and woody vegetation types [*Calvet*, 2000]. The meta-analysis shows relationships between g_m and $D_{\max}$ for low vegetation and between g_m and f_0 for high vegetation. Furthermore, it seems that plants react in two different ways to soil moisture stress. There are plants that try to avoid stress, by reducing the evaporation via stomatal regulation. This stress strategy is typified as defensive. Others apply an offensive strategy in order to resist stress, by a more efficient root water uptake or a more rapid growing cycle. Among species within the seven vegetation classes of CTESSEL often both strategies occur. Therefore, it is not easy to generalize the strategy for each class. It seems most likely that coniferous forest has a defensive strategy, whereas an offensive strategy is assigned to the other classes. Further details on the soil stress parameterization can be found in the work of *Voogt et al.* [2006]. The net CO_2 assimilation calculated at the leaf scale is upscaled at the canopy scale assuming that leaf parameters do not vary within the canopy. The attenuation of the incoming shortwave radiation in the canopy is computed thanks to a simple radiative transfer model [*Calvet et al.*, 1998].

A2. Vegetation Growth Model

[66] The interactive LAI is based on biomass evolution due to photosynthetic activity. The biomass module simulates growth and mortality of the vegetation. The growth of active biomass B is based on the accumulated net CO_2 assimilation over the previous day $A_{n,\text{day}}$ and the LAI is obtained from the biomass following

$$\alpha_B = \frac{B}{\text{LAI}}. \tag{A5}$$

In reality, α_B depends on climate (temperature and CO_2 concentration) and nitrogen fertilization. In order to account for plant morphology, the nitrogen dilution concept is applied for the biomass evolution. The plant N decline model is a well-established agronomical law relating the plant N in nonlimiting N supply conditions to the accumulated aboveground dry matter. The critical plant N is the value of N maximizing growth, and this value decreases for increasing biomass accumulation following a negative power law. The basis of the model is that the metabolic component B of the plant biomass is related to total biomass B_T through an allometric logarithmic law [*Calvet and Soussana*, 2001]. In CTESSEL, the metabolic biomass component is identified as the active biomass and

the relationship between active biomass B and total aboveground biomass B_T is

$$B_T = \left(\frac{B}{c}\right)^{\frac{1}{1-a}}, \tag{A6}$$

where a and c are constant parameters. The total aboveground biomass consists of the active biomass reservoir and the structural aboveground reservoir B_s, which can be considered as the "living" structural biomass, like the stem. For forests, wood is a dead reservoir and does not contribute to B_s. Within the nitrogen dilution model a relationship between the leaf area ratio LAR and the aboveground nitrogen concentration N_T is applied:

$$\text{LAR} = \frac{\text{LAI}}{B_T} = e N_T + f(1.48), \tag{A7}$$

where e and f are called plasticity parameters and are derived per vegetation type. Equation (A7) can be used as a closure equation to estimate α_B

$$\alpha_B = \frac{1}{e N_a + \frac{f}{c B_T^{-a}}}, \tag{A8}$$

where N_a is the nitrogen concentration in the active biomass. It depends on vegetation type and on the nitrogen fertilization. For further details and derivations see *Calvet and Soussana* [2001]. In this way, α_b has become a model variable depending on B_T. However, for global simulations, it is desirable to keep α_b as a constant parameter in order to let α_B represent rather intrinsic plant characteristics denoting a biological adaptation to average climate and growing conditions [*Calvet and Soussana*, 2001]. For that purpose, equation (A8) can only be solved by iteration. Moreover, LAR and N_T data to derive the plasticity parameters by regression is lacking. However, data is available for leaves in the form of the specific leaf area SLA and the nitrogen content in leaves N_L:

$$\text{SLA} = \frac{\text{LAI}}{B_L} = e N_L + f. \tag{A9}$$

Both the iteration issue and the availability of data to derive e and f give rise to modify the nitrogen dilution module. Equation (A8) is simplified by considering α_B as the ratio of the biomass of green leaves to LAI:

$$\alpha_B = \frac{1}{\text{SLA}} = \frac{1}{e N_L + f}. \tag{A10}$$

A3. Soil Respiration and NEE Parameterization

[67] Since the biomass model is not coupled to a soil model, soil respiration needs to be parameterized in another way. In CTESSEL the respiration is calibrated in order to simulate a zero net CO_2 exchange over a multiyear period. The CO_2 ecosystem respiration is split into two terms. The first is the dark respiration R_d (parameterized from A_m). The second respiration term represents both heterotrophic respi-
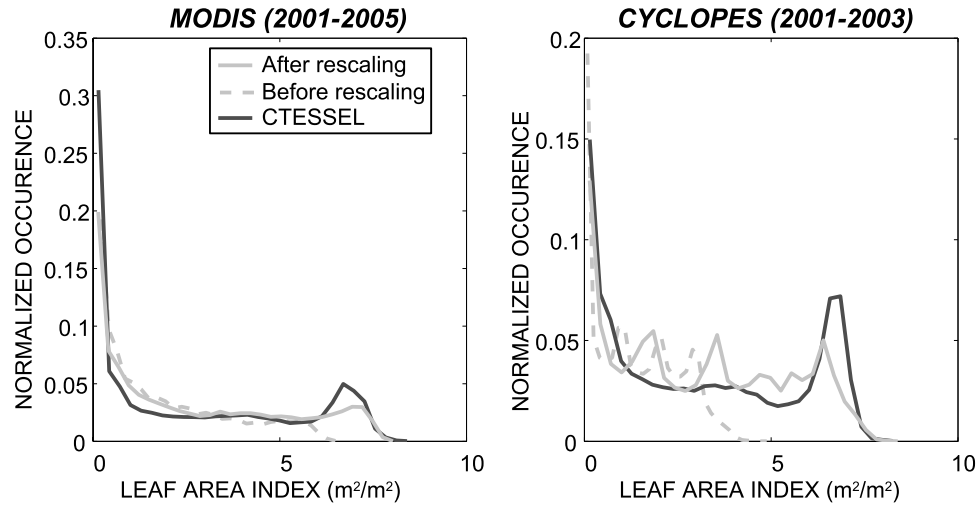
Figure B1. LAI histograms for CYCLOPES and MODIS before and after rescaling to CTESSEL open loop LAI values. The availability of LAI products is 2001–2003 and 2001–2005 for CYCLOPES and MODIS, respectively.

ration from the soil and autotrophic respiration from the above- and belowground structural biomass and named, hereinafter, R_{soilstr}. A Q_{10} function is used for its parameterization:

$$R_{\mathrm{soiltr}} = R_{\mathrm{eco}} - R_d = R_0 Q_{10}^{\left(\frac{T_{\mathrm{soil}}-25}{10}\right)}, \qquad (A11)$$

where R_0 is the reference respiration at 25°C, T_{soil} is the temperature of the second soil layer and Q_{10} is fixed at 2.0. R_0 is determined per vegetation type in each grid box, assuming equilibrium between multiannual net CO_2 assimilation, harvest and residual respiration.

[68] Finally, the net ecosystem CO_2 exchange (NEE) is given by

$$NEE = A_n - R_{\mathrm{soilstr}}. \qquad (A12)$$

Appendix B: Preprocessing of the Satellite LAI Products

[69] The satellite-derived LAI data sets are both negatively biased with regards to the CTESSEL open loop simulations (forced with FORCING1). The reflectances from which the LAI is retrieved are perturbed and noisy. Because of atmospheric and directional effect, the absolute LAI values are subject to question. The phenology (start, end and peak of the growing season) is assumed to be better depicted by the satellite. In addition, the simplified 2DVAR, similarly to others DA methods derived from the Best Linear Unbiased Estimates, aims at correcting for Gaussian errors with a mean equal to zero and not for systematic errors. Therefore, the satellite LAI products are debiased by rescaling their histograms to the CTESSEL LAI simulations by adjusting the 5th and 95th percentiles. As such, there is a risk that part of the removed bias originates from the model and not from the observations. This kind of bias correction would tend to reinforce the potential model bias. To limit

this risk, the rescaling is computed by vegetation types (thanks to the simplified ECOCLIMAP vegetation map) independently of the geographical area and of the time as one can expect that the bias is localized either temporally (during the summer month where the cloud cover is important for instance) or spatially (over area where the vegetation is dense). This approach ensures that the phenology is kept and that only the absolute values are adjusted. Finally, this approach has shown to give better results than by taking all the LAI values independently of their vegetation types (not shown). For each grid point, the dominant vegetation type is determined and the rescaling is applied by vegetation types over the study window. For information, four dominant vegetation types are present: evergreen, grass C3, grass C4 and crops C3. The histograms of CYCLOPES and MODIS LAI products (for all vegetation types) are compared to CTESSEL simulated values before and after rescaling at Figure B1. The rescaling is really efficient for MODIS with a smooth curves over the range of LAI values apart from the higher LAI corresponding to equatorial forest that are not totally retrieved after rescaling. In contrast, the histogram of CYCLOPES LAI is more discontinuous with peaks at 2, 4, and 6 m^2/m^2 which are not reproduced by CTESSEL. After rescaling, the peaks remain and the rescaled histogram does not match the CTESSEL one properly.

[70] The MODIS data set is available over the whole study period (2001–2005) and has LAI values in better agreement with CTESSEL after rescaling. Nevertheless, it shows strong scatter (see time series in Figure 6a). The second step in the preprocessing is to apply a temporal smoother to MODIS LAI products. The CYCLOPES LAI already presents a smooth temporal evolution. This corroborates the results of *Weiss et al.* [2007] who show the smoothness of the CYCLOPES products by comparison to other satellite-derived LAI data sets.

[71] The evolution of the CTESSEL LAI has a relatively smooth dynamic and the model is not supposed to match the high-frequency oscillation of the satellite LAI products that

Table B1. Biases Between the CTESSEL Model and the Satellite-Derived Products: LAI Average Bias, Bias at LAI Peak, Peak Date Bias, and Bias of Beginning of Growth Date Before and After Histogram Matching for the Three Latitudinal Bands

	Average Bias (m^2/m^2)	Peak Bias (m^2/m^2)	Peak Date Bias (days)	Growth Date Bias (days)
CTESSEL-CYCLOPES LAI				
Sahel				
Before	−0.23	−0.52	+23	−66
After	−0.52	−1.21	+23	−63
Savannah				
Before	0.83	1.15	+36	−107
After	−0.29	−0.36	+36	−101
Forest				
Before	4.08	3.69	–	–
After	0.96	0.51	–	–
CTESSEL-MODIS LAI				
Sahel				
Before	−0.26	−0.61	+31	−26
After	−0.21	−0.74	+32	−30
Savannah				
Before	0.27	0.38	+13	−115
After	−0.50	−0.26	+13	−120
Forest				
Before	1.95	1.54	–	–
After	0.35	−0.01	–	–

are attributed to atmospheric perturbations, erroneous clouds masking and differences in the acquisition geometrical configuration. Since these perturbations lead to a negative bias, a upper envelope temporal smoother is applied to the time series over each grid point independently. The Best Index Slope Extraction filter is chosen (BISE filter) [*Viovy et al.*, 1992]. Since this filter is difficult to parameterize, a preliminary study aiming to test every combination of parameters values is carried out in order to find the best tradeoff between smoothing and preservation of peak values. With this optimal parameter sets, the signal is clear and smooth over most regions. Figure 6b shows the time series of satellite products and CTESSEL LAI (same as Figure 6a) after rescaling and smoothing of MODIS LAIs.

[72] Table B1 displays the bias between the satellite-derived LAI and the CTESSEL open loop: average bias (m^2/m^2), the bias at LAI peak (m^2/m^2), the peak date bias (days) and the bias of beginning of growth date (days) before and after histogram matching for the three latitudinal bands. Following *White et al.* [1997], the date of the beginning of growth has been defined as the date t when the difference between LAI(t) and LAI$_{min}$ becomes higher than 30% of annual amplitude. The calculation of phenology dates for forest has been avoided. The bias reduction is not homogeneous over the region of study and the agreement between MODIS and CYCLOPES has been deteriorated over Sahel. Nevertheless whereas the bias between CYCLOPES and CTESSEL has been strongly increased over Sahel, the rescaling process improves the agreement (in terms of bias) over savannah and forest. Concerning MODIS, the average annual bias is improved over Sahel and forest and only slightly deteriorated over savannah (+0.27 and −0.50 m^2/m^2 before and after histogram matching, respectively). The bias at the LAI peak is higher after rescaling for Sahel but improved for savannah and forest. Concerning the phenology,

the date of peak is not modified by the rescaling process and differences in the beginning of growth before and after bias correction does not exceed 6 days on average (i.e., for CYCLOPES over savannah).

References

Andersson, E., E. Hólm, P. Bauer, A. Beljaars, G. A. Kelly, A. P. McNally, A. J. Simmons, J. N. Thépaut, and A. M. Tompkins (2007), Analysis and forecast impact of the main humidity observing systems, *Q. J. R. Meteorol. Soc.*, *133*, 1473–1485.

Arora, V. K. (2002), Modelling vegetation as a dynamic component in soil-vegetation-atmosphere-transfer schemes and hydrological models, *Rev. Geophys.*, *40*(2), 1006, doi:10.1029/2001RG000103.

Balsamo, G., F. Bouyssel, and J. Noilhan (2004), A simplified bi-dimensional variational analysis of soil moisture from screen-level observations in a mesoscale numerical weather-prediction model, *Q. J. R. Meteorol. Soc.*, *130*, 895–915, doi:10.1256/qj.02.215.

Baret, F., et al. (2008), LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION. part 1: Principles of the algorithm, *Remote Sens. Environ.*, *110*(3), 275–286.

Beljaars, A. C. M., and P. Viterbo (1999), *Soil Moisture-Precipitation Interaction: Experience With Two Land Surface Schemes in the ECMWF Model: Global Energy and Water Cycles*, edited by K. Browning and R. Gurney, pp. 223–233, Cambridge Univ. Press, Cambridge, U. K.

Bennett, A. F. (2002), *Inverse Modeling of the Ocean and Atmosphere*, 346 pp., Cambridge Univ. Press, New York.

Calvet, J.-C. (2000), Investigating soil and atmospheric plant water stress using physiological and micrometeorological data, *Agric. Meteorol.*, *103*, 229–247, doi:10.1016/S0168-1923(00)00130-1.

Calvet, J. C., and J. F. Soussana (2001), Modelling CO_2 enrichment effects using an interactive vegetation SVAT scheme, *Agric. For. Meteorol.*, *108*, 129–152, doi:10.1016/S0168-1923(01)00235-0.

Calvet, J. C., J. Noilhan, J.-L. Roujean, P. Bessemoulin, M. Cabelguenne, A. Olioso, and J.-P. Wigneron (1998), An interactive vegetation SVAT model tested against data from six contrasting sites, *Agric. For. Meteorol.*, *92*, 73–95, doi:10.1016/S0168-1923(98)00091-4.

Denning, A. S., G. J. Collatz, C. Zhang, D. A. Randall, J. A. Berry, P. J. Sellers, G. D. Colello, and D. A. Dazlich (1996), Simulations of terrestrial carbon metabolism and atmospheric CO_2 in a general circulation model. part 1: Surface carbon fluxes, *Tellus, Ser. B*, *48*, 543–567.

Drush, M. (2007), Initializing numerical weather prediction models with satellite-derived surface soil moisture: Data assimilation experiments with ECMWF's Integrated Forecast System and the TMI soil moisture data set, *J. Geophys. Res.*, *112*, D03102, doi:10.1029/2006JD007478.

Duchemin, B., B. Berthelot, G. Dedieu, M. Leroy, and P. Maisongrande (2002), Normalisation of directional effects in 10-day global syntheses derived from VEGETATION/SPOT. II—Validation of an operational method on actual data sets, *Remote Sens. Environ.*, *81*, 101–113, doi:10.1016/S0034-4257(01)00337-6.

Falk, U., C. Brümmer, N. Brüggemann, R. Wassmann, and J. Szarzynski (2007), Fluxes of carbon and energy above a natural savannah in Burkina Faso, West Africa, *Geophys. Res. Abstr.*, *9*, 09302.

Gibelin, A., J. Calvet, J. Roujean, L. Jarlan, and S. O. Los (2006), Ability of the land surface model ISBA-A-gs to simulate leaf area index at the global scale: Comparison with satellites products, *J. Geophys. Res.*, *111*, D18102, doi:10.1029/2005JD006691.

Gu, Y., S. Belair, J. F. Mahfouf, and G. Deblonde (2006), Optimal interpolation analysis of leaf area index using MODIS data, *Remote Sens. Environ.*, *104*(3), 283–296, doi:10.1016/j.rse.2006.04.021.

Hanan, N. P., P. Kabat, A. J. Dolman, and J. A. Elbers (1998), Photosynthesis and carbon balance of a Sahelian fallow savanna, *Global Change Biol.*, *4*, 523–538, doi:10.1046/j.1365-2486.1998.t01-1-00126.x.

Hiernaux, P. (1984), Distribution des pluies et production herbacée au Sahel: Une méthode empirique pour caractériser la distribution des précipitations journalières et ses effets sur la production herbacée—Premiers résultats acquis dans le Sahel malien, doc. prog. Zones Arides et Semi-Arides, CIPEA Bamako et ILCA Addis Ab eba, Ethiopia.

Jacobs, C. M. J. (1994), Direct impact of atmospheric CO_2 enrichment on regional transpiration, PhD thesis, Wageningen Agric. Univ., Wageningen, Netherlands.

Kabat, P., A. J. Dolman, and J. A. Elbers (1997), Evaporation, sensible heat and canopy conductance of fallow savannah and patterned woodland in the Sahel, *J. Hydrol. Amsterdam*, *188/189*, 494–515, doi:10.1016/S0022-1694(96)03190-3.

Kimes, D. S., Y. Knyazikhin, J. L. Privette, A. A. Abuelgasim, and F. Gao (2000), Inversion methods for physically based models, *Remote Sens. Rev.*, *18*, 381–439.

Knorr, W., and V. Lakshmi (2001), Assimilating fAPAR and surface temperature into a land surface and vegetation model, in *Land Surface Hydrology and Climate: Observations and Modeling, Water Sci. Appl. Ser.*, vol. 3, edited by V. Lakshmi, J. Albertson, and J. Shaake, pp. 177–200, AGU, Washington D. C.

Kucukkaraca, E., and M. Fisher (2006), Use of analysis ensembles in estimating flow-dependent background error variances, *Tech. Mem. 492*, 16 pp., Eur. Cent. for Medium-Range Weather Forecasts, Reading, U. K.

Lacaze, R. (2004), Algorithm theoretical basis document: Customization for LAI, fAPAR, fcover and albedo, report, CYCLOPES Project, Toulouse, France.

Lafont, S., M. Voogt, A. Beljaars, B. J. J. M. Van den Hurk, L. Jarlan, and G. Balsamo (2007), Description and evaluation of the carbon version of TESSEL: CTESSEL model, technical memorandum, 37 pp., Eur. Cent. for Medium-Range Weather Forecasts, Reading, U. K.

Le Houérou, H. N. (1989), *The Grazing Land Ecosystems of the African Sahel, Ecol. Stud.*, vol. 75, 282 pp., Springer, Berlin.

Leroux, M. (1972), Climatologie dynamique de l'ouest africain, in *Atlas International de l'Ouest Africain*, plates 14–15, Inst. Fondamental d'Afr. Noire, Univ. de Dakar, Senegal.

Lo Seen, D., E. Mougin, S. Rambal, A. Gaston, and P. Hiernaux (1995), A regional Sahelian grassland model to be coupled with multispectral satellite data. II. Towards its control by remotely sensed indices, *Remote Sens. Environ.*, *52*, 194–206, doi:10.1016/00340-4257(94)00127-9.

Masson, V., J. L. Champeaux, F. Chauvin, C. Mériguet, and R. Lacaze (2003), A global database of land surface parameters at 1km resolution for use in meteorological and climate models, *J. Clim.*, *16*, 1261–1282.

Moulin, S., L. Kergoat, N. Viovy, and G. Dedieu (1997), Global-scale assessment of vegetation phenology using NOAA/AVHRR satellite measurements, *J. Clim.*, *10*, 1154–1170.

Muñoz Sabater, J., L. Jarlan, J. C. Calvet, F. Bouyssel, and P. De Rosnay (2007), From near surface to root zone soil moisture using different assimilation techniques, *J. Hydrometeorol.*, *8*, 194–206, doi:10.1175/JHM571.1.

Muñoz Sabater, J. L., et al. (2008), Joint assimilation of surface soil moisture and LAI observations using a simplified 1DVAR: The smosrex case study, *Agric. For. Meteorol.*, *148*(8–9), 1362–1373.

Myneni, R. B., R. R. Nemani, and S. W. Running (1997), Estimation of global leaf area index and absorbed par using radiative transfer models, *IEEE Trans. Geosci. Remote Sens.*, *35*, 1380–1393, doi:10.1109/36.649788.

Philippon, N., and B. Fontaine (2001), The relationship between the Sahelian and previous second Guinean rainy seasons: A monsoon regulation by soil wetness, *Ann. Geophys.*, *20*, 575–582.

Prince, S. (1991), Satellite remote sensing of primary production: Comparison of results for Sahelian grasslands 1981–1988, *Int. J. Remote Sens.*, *12*, 1301–1311, doi:10.1080/01431169108929727.

Randerson, J. T., M. V. Thompson, T. J. Conway, I. Y. Fung, and C. B. Field (1997), The contribution of terrestrial sources and sinks to trends in the seasonal cycle of atmospheric carbon dioxide, *Global Biogeochem. Cycles*, *11*, 535–560, doi:10.1029/97GB02268.

Rayner, P. J., M. Scholze, W. Knorr, T. Kaminski, R. Giering, and H. Widmann (2005), Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS), *Global Biogeochem. Cycles*, *19*, GB2026, doi:10.1029/2004GB002254.

Reichle, R. H., R. D. Koster, P. Liu, S. P. P. Mahanama, E. G. Njoku, and M. Owe (2007), Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR), *J. Geophys. Res.*, *112*, D09108, doi:10.1029/2006JD008033.

Roujean, J.-L., and R. Lacaze (2002), Global mapping of vegetation parameters from POLDER multiangular measurements for studies of surface-atmosphere interactions: A pragmatic method and its validation, *J. Geophys. Res.*, *107*(D12), 4150, doi:10.1029/2001JD000751.

Shukla, J., and Y. Mintz (1982), Influence of land-surface evapotranspiration on the Earth's climate, *Science*, *215*, 1498–1501, doi:10.1126/science.215.4539.1498.

Sivakumar, M. V. K. (1990), Exploiting rainy season potential from the onset of rains in the Sahelian zone of West Africa, *Agric. For. Meteorol.*, *51*, 321–332, doi:10.1016/0168-1923(90)90116-N.

Tian, Y., Y. Zhang, Y. Knyazikhin, R. B. Myneni, and S. W. Running (2000), Prototyping of MODIS LAI/FPAR algorithm with LASUR and Landsat data, *IEEE Trans. Geosci. Remote Sens.*, *38*, 2387–2401, doi:10.1109/36.868894.

Tucker, C. J., C. L. Vanpraet, M. J. Sharman, and G. Van Ittersum (1985), Satellite remote sensing of total herbaceous biomass production in the Senegalese Sahel: 1980–1984, *Remote Sens. Environ.*, *17*, 233–249, doi:10.1016/0034-4257(85)90097-5.

Tucker, C. J., Y. I. Fung, C. D. Keeling, and R. H. Gammon (1986), Relationship between atmospheric CO_2 variations and a satellite derived vegetation index, *Nature*, *319*, 195–199, doi:10.1038/319195a0.

Tucker, C. J., J. E. Pinzon, M. E. Brown, D. Slayback, E. W. Pak, R. Mahoney, E. Vermote, and N. El Saleous (2005), An extended AVHRR 8-km NDVI data set compatible with MODIS and SPOT vegetation NDVI data, *Int. J. Remote Sens.*, *26*, 4485–4498, doi:10.1080/01431160500168686.

Van den Hurk, B. J. J. M., P. Viterbo, A. C. M. Beljaars, and A. K. Betts (2000), Offline validation of the ERA40 surface scheme, *Tech. Memo 295*, 43 pp., Eur. Cent. for Medium-Range Weather Forecasts, Reading, U.K.

Van den Hurk, B. J. J. M., P. Viterbo, and S. O. Los (2003), Impact of leaf area index seasonality on the annual land surface evaporation in a global circulation model, *J. Geophys. Res.*, *108*(D6), 4191, doi:10.1029/2002JD002846.

Veenendaal, E. M., O. Kole, and J. Lloyd (2004), Seasonal variation in energy fluxes and carbon dioxide exchange for a broad-leaved semi-arid savanna (Mopane woodland) in Southern Africa, *Global Change Biol.*, *10*, 318–328, doi:10.1111/j.1365-2486.2003.00699.x.

Verhoef, A., S. J. Allen, H. A. R. De Bruin, C. M. J. Jacobs, and B. G. Heusinkveld (1996), Fluxes of carbon dioxide and water vapour from a Sahelian Savanna, *Agric. For. Meteorol.*, *80*, 231–248, doi:10.1016/0168-1923(95)02294-5.

Viovy, N., O. Arino, and A. S. Belward (1992), The Best Index Slope Extraction (BISE)—A method for reducing noise in NDVI time series, *Int. J. Remote Sens.*, *13*, 1585–1590, doi:10.1080/01431169208904212.

Voogt, M. H., B. J. J. M. van den Hurk, and C. M. J. Jacobs (2006), The ECMWF land surface scheme extended with a photosynthesis and LAI module tested for a coniferous site, *WR-06-02*, 22 pp., R. Neth. Meteorol. Inst. (KNMI), De Bilt, Netherlands.

Walker, J. P., and P. R. Houser (2001), A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations, *J. Geophys. Res.*, *106*, 11,761–11,774.

Wallace, J. S., and C. J. Holwill (1997), Soil evaporation from Tiger-bush in south-west Niger, *J. Hydrol. Amsterdam*, *188–189*, 426–442, doi:10.1016/S0022-1694(96)03185-X.

Weiss, M., F. Baret, G. J. Smith, I. Jonckheere, and P. Coppin (2004), Review of methods for in situ leaf area index (LAI) determination: part II. Estimation of LAI, errors and sampling, *Agric. For. Meteorol.*, *121*(1–2), 37–53, doi:10.1016/j.agrformet.2003.08.001.

Weiss, M., F. Baret, S. Garrigues, and R. Lacaze (2007), LAI and FAPAR CYCLOPES global products derived from vegetation. part 2: Validation and comparison with MODIS C4 products, *Remote Sens. Environ.*, *110*, 317–331, doi:10.1016/j.rse.2007.03.001.

White, M. A., P. E. Thornton, and S. W. Running (1997), A continental phenology model for monitoring vegetation responses to interannual climatic variability, *Global Biogeochem. Cycles*, *11*, 217–234, doi:10.1029/97GB00330.

Zeng, N., J. D. Neelin, K.-M. Lau, and C. J. Tucker (1999), Enhancement of interdecadal climate variability in the Sahel by vegetation interaction, *Science*, *286*, 1537–1540, doi:10.1126/science.286.5444.1537.

———————————

G. Balsamo, A. Beljaars, L. Jarlan, and S. Lafont, ECMWF, Shinfield Park, Reading RG2 9AX, UK. (lionel.jarlan@ecmwf.int)

J. C. Calvet, Météo-France, Centre National de Recherches Météorologiques/GMME-MC2, 42 avenue Coriolis, F-31057 Toulouse CEDEX 1, France.

E. Mougin, CESBIO, 18 avenue Edouard Belin bpi 2801, F-31401 Toulouse CEDEX 9, France.