### Variational Assimilation of Land Surface Temperature within the ORCHIDEE Continental Surface model

**H. S. Benavides Pinjosovsky1 2 3, C. Ottlé2, S. Thiria1, P. Maugis2, J. Brajard1, and F. Bradran1**

[1]{Laboratoire d’Océanographie et du Climat: Expérimentations et approches numériques, IPSL Paris, France}

[2]{Laboratoire des Sciences du Climat et de l’Environnement, IPSL, CNRS-CEA-UVSQ, Gif-sur-Yvette, France}

[3]{CLIMMOD Engineering}

1. Correspondence to: H. S. Benavides Pinjosovsky ([hbplod@locean-ipsl.upmc.fr](mailto:hbplod@locean-ipsl.upmc.fr)) and C. Ottlé (catherine.ottle@lsce.ipsl.fr)

# Abstract

1. *The energy and water budget modules of the ORCHIDEE land surface model (SECHIBA) describes the exchanges of water and energy between the surface and the atmosphere. In the present paper, the adjoint semi-generator software denoted YAO was used as a framework to implement a 4D-VAR assimilation method. The objective of this work is to deliver the adjoint model of SECHIBA (SECHIBA-YAO) obtained with YAO to provide an opportunity for scientists to perform their own assimilation. SECHIBA-YAO allows the control of the eleven most influent internal parameters of SECHIBA or of the* *initial condition of the soil water content by observing the land surface temperature or the brightness temperature. At a given location, soil and climate conditions, we can use remote sensed data in the assimilation process. The paper presents the fundamental principles of the 4D-Var assimilation, the semi-generator software YAO and some experiments showing the accuracy of the adjoint code distributed.*
2. *Keywords: Sensitivity Analysis, Data Assimilation, Adjoint model, Land Surface Temperature*

# 1. Introduction

Land surface models (LSM) simulate the interactions between the atmosphere and the land surface, which directly influence the exchange of water, energy and carbon with the atmosphere. They are important tools for understanding the main interaction and feedback processes simulating the present climate and making predictions of future climate evolution (Harrison et al., 2009). Such predictions are subject to considerable uncertainties, related to the difficulty to model the highly complex physics with a limited set of equations that does not account for all the interacting processes (Pipunic et al., 2008, Ghent et al. 2011). Understanding these uncertainties is important in order to obtain more realistic simulations.

The main challenge of each dynamical model, regardless its nature, is to have the appropriate source of information to produce an accurate response. Observations sample the system of interest in space and time. These measurements provide essential information on the model dynamics and contribute to the understanding of the system evolution (Lahoz et al. 2010). Data assimilation adds observations to the model, constraining it to represent the trajectory of the modeled phenomena more accurately. The idea is to merge the measurements with the dynamical model in order to obtain a more accurate estimate of the current and future states of the system, together with the uncertainty estimates in the model states. Two basic methodologies can be used to come up with uncertainties. The sequential approach (Evensen 2003), based on the statistical estimation theory of the Kalman filter, and the variational approach, the so-called 4DVAR (Le Dimet et al., 1986), built from the optimal control theory (Robert et al, 2007). It is well known that both approaches provide the same solution at the end of the assimilation period, for perfect and linear models. But both approaches become very different when the processes under study are highly non linear. The main advantage of 4DVAR comes from its integration in time achieved during the assimilation of the observations, giving rise to a global trajectory of the model optimized over the assimilation time window.

Variational data assimilation has been widely used in land surface applications. The assimilation of land surface temperature (LST) is suitable for an extensive range of environmental problems. As mentioned in Ridler et al. (2012), LST is an excellent candidate for model optimization since it is solution of the coupled energy and water budgets, and permits to constrain parameters related to evapotranspiration and indirectly to soil water content. In Castelli et al. (1999), a variational data assimilation approach is used to include surface energy balance in the estimation procedure as a physical constraint (based on adjoint techniques). The authors worked with satellite data, and directly assimilated soil skin temperatures. They conclude that constraining the model with such observation improves model flux estimates, with respect to available measurements. In Huang et al. (2003) the authors developed a one-dimensional land data assimilation scheme based on an ensemble Kalman filter, used to improve the estimation of land surface temperature profile. They demonstrate that the assimilation of LST into land surface models is a practical and effective way to improve the estimation of land surface state variables and fluxes. Reichle et al. (2010) performs the assimilation of satellite-derived skin temperature observations using an ensemble-based, ofﬂine land data assimilation system. Results suggest that retrieved fluxes provide modest but statistically significant improvements. However, these authors noted strong biases between LST estimates from *in situ* observations, land modeling, and satellite retrievals that vary with season and time of the day. They highlighted the importance of taking these biases into account. Otherwise large errors in surface ﬂux estimates can result. Ghent et al. (2011) investigated the impacts of data assimilation on terrestrial feedbacks of the climate system. Assimilation of LST helped to constrain simulations of soil moisture and surface heat fluxes. Ridler et al. (2012), tested the effectiveness of using satellite estimates of radiometric surface temperatures and surface soil moisture to calibrate a Soil–Vegetation–Atmosphere Transfer (SVAT) model, based on error minimization of temperature and soil moisture model outputs. Flux simulations were improved when the model is calibrated against *in situ* surface temperature and surface soil moisture versus satellite estimates of the same fluxes. In Bateni et al. (2013), the full heat diffusion equation is employed in the variational data assimilation scheme as an adjoint (constraint). Deviations terms of the evaporation fraction and a scale coefficient are added as penalization terms in the cost function. Weak constraint is applied to data assimilation with model uncertainty, accounting in this way for model error. The cost function associated with this experiment contains a term that penalizes the deviation from prior values. When assimilating LST into the model, the authors proved that the heat diffusion coefficients are strongly sensitive to specific deep land surface temperature. As a conclusion, it can be seen that the assimilation of LST can improve the model simulated flows.

In the present study, we focus on part of the ORCHIDEE Land Surface Model denoted SECHIBA (Ducoudré et al. 1993) and demonstrate the ability of 4DVAR to estimate a set of its inner parametersas well as initial conditions of surface soil water contentby observing the brightness temperature or the soil temperature. Dedicated software (denoted SECHIBA-YAO) is made available to enable the reader to experiment by himself with his data. The development of SECHIBA-YAO was conducted by using the adjoint semi-generator software denoted YAO developed at LOCEAN-IPSL (Nardi et al. 2009). YAO serves as a framework to design and implement dynamic models, helping to generate the adjoint of the model which permits to compute the model gradients. SECHIBA-YAO provides an opportunity for scientists to control the ten most influent internal parameters of SECHIBA by observing the land surface temperature or the brightness temperature. For a given location, soil and climate conditions, one ran twin experiments and also assimilation with remote sensed data. The twin experiments conducted on actual sites were used to demonstrate the accuracy and usefulness of the code and the potential of 4D-VAR when dealing with LST assimilation.

The paper is structured as follows. In Section 2 model and data used to illustrate the capabilities of the SECHIBA-YAO are detailed. In Section 3, fundamentals of variational data assimilation are presented. In addition, principles of YAO and of its associated modular graph formalism are exposed. The principle of the computation of the adjoint with YAO is provided. The implementation of SECHIBA-YAO and the details of the experiments that prove the efficiency of the 4D-Var assimilation are also subject of Section 3. Sensitivity experiments and simple twin experiments at a single location are presented in Section 4. These experiments illustrate the convenience of YAO to optimize control parameters. Finally, the specificities of the distributed software are given in Section 5 with a demo script that illustrates the properties of the software.

## 2. Models and Data

ORCHIDEE is a Land Surface Model developed at the “Institut Pierre Simon Laplace (IPSL)” in France. ORCHIDEE is a mechanistic dynamic global vegetation model (Krinner et al., 2005) representing the continental biosphere and its different processes. It is part of the IPSL (Institut Pierre Simon Laplace) earth system model (LMDZ, Hourdin et al., 2006; Dufresne et al., 2013) and is composed of 3 modules: SECHIBA, STOMATE and LPJ. SECHIBA computes the water and energy budgets at the biosphere-atmosphere interface, as well as the Gross Primary Production (GPP); STOMATE (Friedlingstein et al., 1999) is a biogeochemical model which represents the processes related to the carbon cycle, such as carbon dynamics, the allocation of photosynthesis respiration and growth maintenance, heterotrophic respiration and phenology and finally, LPJ (Sitch et al., 2003) models the global dynamics of the vegetation, interspecific competition for sunlight as well as fire occurrence. ORCHIDEE has different time scales: 30-minutes for energy and matter, 1-day for carbon processes and 1-year for species competition processes. The full description of ORCHIDEE can be consulted in Ducoudré et al., 1993, Krinner et al., 2005, d’Orgeval et al., 2006, Kuppel et al., 2012. In the present study, ORCHIDEE 1.9 version is used in a grid-point mode (one given location), forced by the corresponding local half-hourly gap-filled meteorological measurements obtained at the flux towers. In this study, only the SECHIBA module is activated.

SECHIBA (Schématisation des Echanges Hydriques à l'Interface Biosphère-Atmosphère) (Ducoudré et al., 1993) is a land surface model. It solves every half hour the energy budget of the surface and the soil water budget. The land surface is represented as a whole system composed of various fractions of vegetation types called PFT (Plant Functional Type). A single energy budget is performed for each grid point, but water budget is calculated for each PFT fraction. The resulting energy and water fluxes between atmosphere, ground and the retrieved temperature represent the canopy ensemble and the soil surface. The main fluxes modeled are the net radiation (*Rn*), soil heat flux (*Q*), sensible (*H*) and latent heat (*LE*) fluxes between the atmosphere and the biosphere, land surface temperature (*LST*) and the soil water reservoir contents. Energy balance is solved once, with a subdivision only for *LE* in bare soil evaporation, interception and transpiration for each type of vegetation. Water balance is computed for each fraction of vegetation (Plant Functional Type or PFT) present in the grid. The SECHIBA version used in this work models the hydrological budget based on a two-layer soil profile (Choisnel, 1977). The two soil layers represent respectively the surface and the total rooting zone. The soil is considered homogeneous with no sub-grid variability and of a total depth of *htot = 2m.* The soil bottom layer acts like a bucket that is filled with water from the top layer. The soil is filled from top to bottom with precipitation; when evapotranspiration is higher than precipitation, water is removed from the upper reservoir. Runoff arises when the soil is saturated. SECHIBA inputs are: *Rlw* the incoming infrared radiation; *Rsw* the incoming solar radiation; *P*the total precipitation (rain and snow); *Ta* the air temperature; *Qa* the air humidity; *Ps* the atmospheric pressure at the surface and *U* the wind speed.

In SECHIBA-YAO, the simulated LST is hemispheric and does not account for solar configuration and viewing angle effects, it is also integrated on all the solar spectrum. In order to compute brightness temperature from LST, there is a direct link between both variables. Neglecting the directional effects, the total energy emitted by the surface (Rad) can be computed using the following expression (Ducoudré et al. 1993)

 (Eq 1)

In this equation,  is the surface emissivity,  is the multiplicative factor for emissivity and  is the long wave incident radiation that is an input forcing of SECHIBA. Svendsen et all (1990) proposed a transfer function to link the surface emitted radiance towards an observed brightness temperature *TB* measured in the [8,14] spectral band The empirical formulation is given by the expression

 (Eq 2)

This formulation was added in SECHIBA-YAO to allow the calculation of *TB* from satellite observations.

In the following the capabilities of the 4D-VAR is demonstrated in a series of assimilation experiment using the data provided by the FLUXNET network. SECHIBA-YAO can be run using other data as long as the inputs needed to operate SECHIBA are completed. FLUXNET (Baldocchi et al., 2001) is a network coordinating regional and global analysis of observations from micrometeorological tower sites. The flux tower sites use eddy covariance methods (Aubinet et al. 2012) to measure the exchange of carbon dioxide (CO2), water vapor, and energy between terrestrial ecosystems and the atmosphere.

Measurement towers sprang up around the world, grouped in regional networks. The data from all networks is accessible to the scientific community via the Fluxnet website (<http://www.fluxdata.org>). In this work, we selected 2 sites: Harvard Forest and Skukuza Kruger National Park; both present contrasted climate and land surface properties suitable to test the tools developed and assess model parameters sensitivities. Only climate measurements with the same sampling frequency (30 minutes) from both sites are used to force SECHIBA. Vegetation characteristics are prescribed and only homogeneous grids are considered. Two cases were studied with grassland (PFT 11) and bare soil (PFT 0).

*Skukuza Kruger National Park*

Located in South Africa at 25° 1' 11" S and 31° 29' 48" E, , this Fluxnet site was established in 2000. The tower overlaps two distinct savanna types and collects information about land-atmosphere interactions. The climate is Subtropical-Mediterranean. The total mean annual precipitation is 650 mm, with an altitude of 150 m and the mean annual temperature is 22.15 ºC.

*Harvard Forest*

Located in the United States of America, on land owned by Harvard University, the station is located at 42º53'78'' N and 72º17'15'' W. It was established in 1991. The site has a Temperate-Continental climate with hot or warm summers and cold winters. The annual mean precipitation is 1071 mm, the mean annual temperature is 6.62 ºC and the altitude is 340 m.

# 3. The Methodology

# 3.1 Variational assimilation

Variational assimilation (4D-VAR) (Le Dimet et al. 1986) considers a physical phenomenon described in space and its time evolution. It thus requires the knowledge of a direct dynamical model *M*, which describes the time evolution of the physical phenomenon. *M* allows connecting the geophysical variables studied with observations. By varying some geophysical variables (control variables); assimilation seeks to infer the physical variables that led to the observation values. These physical variables can be, for example, initial conditions or parameters of *M*.

The basic idea is to determine the minimum of a cost function ***J*** that measures the misfits between the observations and the model estimations. Due to the complexity of this function, the desired minimum is classically obtained by using gradient methods, which implies the use of the adjoint model of *M*. This model is derived from the equations of the direct model *M*. The adjoint model estimates changes in the control variables in response to a disturbance of the output values calculated by *M*. It is therefore necessary to proceed in the backward direction to the direct model calculations, which means to use the transpose of the Jacobean matrix with respect to the control parameters. When observations are available, the adjoint allows minimizing the cost function ***J***.

Formalism and notations for variational data assimilation are taken from Ide et al., (1997). *M* represents the direct model, **x**(t0) is the initial state of the model and **k** represents the vector of the inner model parameters to be controlled, so **x***(ti)* = *Mi*(**k**, **x**(*t*0)), where *Mi*(**k**, **x**(*t*0)) is represented by . The tangent linear model is noted **M**(*ti*, *ti*+1), which is the Jacobean matrix of **M**, in **x**(ti). The adjoint model  is the linear tangent transpose, defined as:



Eq(3)



**M** is used to estimate variables, which are most often observed from an observation operator **H**, permitting to compare the observed values **y0** with respect to the **y** calculated by the composition **H°M**, when they are available. The cost function *J* will be defined in terms of observations, so **Hi** allows us to estimate the variables **yi**, from the state vector **x(ti*).*** We suppose that where *i* is a random variable with zero mean. This term represents the sum of the model, observation and scaling error. Finally, the most general form of the cost function is defined as follows:



Eq (4)



The background vector is defined as **k*b***, which is an *a priori* state vector. The first part of the cost function represents the discrepancy to **k**b and acts as a regularization term. The second part represents the distance between the observations and the model estimates. **B** is the covariance error matrix of **k**b and **R**i is the covariance error matrix of **y**o at time ti. The objective of this work is to show the capacity of 4DVAR to help determining the value of the principal inner parameters **k** of SECHIBA and the initial conditions for Surface Water Content. The distributed software allows the reader to do its own experiments using synthetic or actual data. When the observations are synthetic (produced by the model itself) no transfer function from the estimation to the observation are needed, and **H** is taken as the identity matrix. If actual data are used, a specific **H** is used that transforms the soil temperature into brightness temperature (see section Model and Data).

The minimization of the cost function (Eq 4) is based on gradient-descent approaches. The cost function gradient has the form Eq (5)



where and are the gradients of the cost function *J* with respect to **k** and **y*i*** respectively.



The expression above allows us to computeby knowing, in the form of a matrix product of this term by the matrix , corresponding to the transpose of the Jacobian Matrix. The development of calculation gives the expression of the gradient of **y** (equation 2):



Eq (6)



The Control parameters are adjusted several times until a stopping criterion is reached. The iterations of the gradient method allow us to approach the desired solution, in order to satisfy a stopping criterion that could be, for example, a certain threshold on the norm of the cost function gradient.

## 3.2 YAO

Variational data assimilation requires the computation of the adjoint code of the direct model, which is a heavy and complex task, especially for a large model such as SECHIBA. Usually, the adjoint code is computed with the help of specific softwares (automatic differentiators) (e.g., Bischof et al.,1996; Giering and Kaminski, 2003; Hascoët and Pascual, 2004). These softwares are appropriate for the differentiation of large codes, but their use will be optimal only under specific coding conventions and a good level of modularity of the codes (Talagrand, 1991). Moreover, manual optimization of the produced code is often necessary. Therefore, in many practical cases the automatic production of code will not be totally optimal in terms of flexibility (e.g., when the direct model is updated frequently, one has to re-differentiate the whole code). These considerations motivated the development of a slightly different but complementary approach that focuses on the high-level structure of the numerical models, embedding implementation details inside simple entities that can be easily updated. This has led to the development of the YAO assimilation software at LOCEAN/IPSL (<https://skyros.locean-ipsl.upmc.fr/~yao/>). YAO is based on the decomposition of a numerical model into elementary modules interconnected by directional links. On one side, the structure of the model (variables, dependencies...) is described as a graph structure. On the other side, the details of the physics are coded inside C/C++ basic modules that are ideally simple. The user can therefore separate the “high-level” structure of the model from implementation details. It is also very easy to update a numerical code within this framework. Regarding the assimilation strategy, YAO computes the tangent linear and adjoint codes from the elementary jacobians of each module (provided by the user). Adjoint/cost function test tools are also available. Finally, YAO includes routines devoted to classical assimilation scenarii (incremental form ...) and is interfaced with the M1QN3 minimizer (Gilbert and Lemaréchal, 1989).

**Graph formalism**

In YAO, a numerical model must be described as an ensemble of modules related by connections in order to form a graph. Let us define more precisely the main components of the graph:

* a ***module*** is a basic entity of computation, representing a deterministic (but possibly nonlinear) function transforming an input vector into an output vector. A module is viewed graphically as a node of the graph, the sizes of the vectors correspond to the number of input and output connections associated with the node.
* a ***basic connection*** is an oriented link relating two nodes of the graph. Most basic connections usually represent the transmission of the output of one module taken as input by another one.
* The external context is the ensemble of data input and output points used as external data by a whole graph at a specific level of abstraction. Basic connections can move from a data input point located in the external context towards one or several module(s) (for instance modules needing the specification of some initial conditions, boundary conditions or model parameters). Inversely, the global outputs of the model leave from a module towards a data output point located in the external context.
* The modular graph is the ensemble of the modules and of their connections. It must be acyclic so that a topological order may be defined on the nodes of the graph (i.e., if there is connection *Fp* → *Fq*, then *Fp* should be computed before *Fq*) (see Figure 1)

Typically, a modular graph must describe the equations governing the system of interest and each physical variable appearing in the governing equations must be associated with a specific module. However, supplementary modules can also be defined to represent temporary variables required to simplify computations for complex equations. The user has generally to specify modules at a single point (*i*, *j*, *k*, *t*) of space (*i*, *j*, *k*) and time (*t*), and the names and space-time locations (e.g. *i*+1, *j*-1, *k*, *t*-1) of the discretized variables taken as inputs. From the local description of the equations, YAO is able to build a model on a given space domain and on a given number of time steps by automatically replicating the local graph in space-time (cf. Figure 2)).

By passing the different modules in topological order, YAO is clearly able to emulate the global model and to calculate the global model outputs given model initial conditions and parameters.

Now, we will see that the usefulness of the graph modular approach is reinforced when the jacobian matrix of each basic function is known. For a basic function F such that y = F( x ), the jacobian matrix F relates a perturbation of the inputs to the associated perturbation of outputs: **dy** = **F dx**. Since the jacobian of a composition of functions is the product of the elementary jacobians, the tangent linear model associated with a modular graph may also be obtained by passing the graph in the same topological order.

The “lin-forward” algorithm is the following:

1) Initialize the external context's data input points with a perturbation **dx**i (around a given linearization point)

2) Pass the modules in topological order and propagate the perturbation

3) Estimate the perturbation output **dy** on output data points in the external context of the graph.

From this procedure, YAO can emulate the global tangent-linear model from elementary jacobians. In the same manner, a backward algorithm may be defined for adjoint computations. From (Eq. 1), it may be shown that the global adjoint will be retrieved by backpropagating the graph, with a few adjustments not detailed here (see, Nardi et al., 2009 for more details on the “backward” algorithm). This property is the basis of the semi-automatic adjoint computation by YAO.

An implementation of a variational assimilation procedure with YAO follows the structure represented in Fig. 3. The YAO compiler builds an executable file following the scheme presented in Fig.3. This file is independent of the assimilation instructions. The executable file reads these instructions when the user calls them. However, it is not compulsory to use an instruction file since YAO accepts a command-line instruction if no instruction file is provided. Due the graph structure of the model and of its adjoint, it is easy to modify the model and its adjoint, e.g. by updating some adequate modules; one can systematically obtain the update global direct model and the global adjoint

As mentioned in the introduction, this paper only gives access to a compiled version of SECHIBA-YAO and allows only to perform some assimilation experiments related to the control of the ten most influent internal parameters of SECHIBA by observing the land surface temperature or the brightness temperature using equation (Eq1 and Eq 2)). YAO is a free software that gives the opportunity to modify the SECHIBA code provided in this paper.

## 3.3 Development of SECHIBA-YAO

The implementation of SECHIBA in YAO starts with the definition of the modular graph describing the dynamics of the model (see ANNEX A). Elementary processes and interconnections between modules are defined in order to catch the essence of the model. The modular graph is the basis of all the integration processes made by YAO. Direct and adjoint models are computed following the modular graph structure. The modular graph was built as follows:

* Every component of the original code was carefully studied line by line directly.
* A list of inputs and outputs for each subroutine was made, for every routine of SECHIBA. This permits to exactly knows the information flow in the model.
* A second zoom in the subroutines was made in order to understand the internal dynamics of the code. This is the last step in the modular graph definition. When studying the subroutines, they were very general and a division into simpler elements was inevitable, with the purpose of reducing the coupling and increasing the cohesion of the modules. The idea is to have a scalable code. Uncoupled modules give more independence when changing part of the model. Cohesive modules help to understand the model.
* The initial six subroutines in the SECHIBA-Fortran code are split into 130 modules by the SECHIBA-YAO modular graph, corresponding to every process modeled by SECHIBA and to a number of transitional modules serving as auxiliary computing.
* It is important to mention that every variable and subroutine name was kept as in the original model. If a user or developer of SECHIBA-Fortran sees the implementation in YAO, he will find his way easily.

### 3.3.1 Direct model

After defining the modular graph in YAO, the second step in the SECHIBA-YAO implementation is the coding of the direct and the derivatives of the modules. This consists in coding the different modules directly with YAO meta-language. Every module is represented as a script and the different processes attributed to the module are implemented inside the script, allowing a better control of the physics, i.e. any change in the physics could be made easily. In SECHIBA-YAO, the second approach was used.

### 3.3.2 Module Derivatives

Once the direct model has been coded and validated, there are two options to code the derivatives: they can be coded line-by-line based on the forward computing, in order to obtain the Jacobian matrix of the module, or they can also be produced routinely, using an automatic differentiation tool (for example, Tapenade ([Hascoët](http://hal.archives-ouvertes.fr/index.php?action_todo=search&s_type=advanced&submit=1&search_without_file=YES&f_0=AUTHORID&p_0=is_exactly&halsid=4rs19qtmt3rd2pnbatvr39pml0&v_0=98550) et al, 2012)). For SECHIBA-YAO, the derivative process is made line-by-line. The outputs are derived with respect to every input. YAO generates automatically, based on these derivatives, the tangent linear and the adjoint of the model.

Nevertheless, the derivative process introduced errors related to the coding process, to inexact derivatives, expressions that were not differentiated among others. In order to reduce it to a minimum number of bugs, the adjoint of the model was validated (as it was made with the direct model). This guarantees the accuracy when performing assimilation. The validation of the adjoint model is presented in section 4. More validations of the direct and the adjoint models are available in Benavides, 2014.

## 4 Data assimilation experiments

1. In this section we present several experiments that have been realized using the SECHIBA-YAO code. They are related to *the control of the eleven most influent internal parameters of SECHIBA or the control of the* *initial condition of the soil water content by observing the land surface temperature or the brightness temperature.*
2. There are two groups of parameters: inner parameters and multiplying factors. The first group corresponds to physical parameters. The second group collects parameters weighting some physical processes of SECHIBA. In the initial model, they all have the same value of 1 indicating that no weights are used, thus the effect of the assimilation is to allow a local adaptation of these weighting factors. The model inner parameters are the following: *rsolcste* is a numerical constant involved in the soil resistance to evaporation. This parameter limits the soil evaporation, so the greater its value the lower the evaporation; *humcste*, *mxeau* and *mindrain* are related to soil water processes, the higher their values, the more water will be available in the model reservoir, affecting water transfers and especially evapotranspiration; *dpucste* represents the soil depth in meters. The other parameters are multiplicative factors. We have *krveg* which is used in the calculation of the stomata resistance, this variable limits the transpiration capacity of leaves, the greater its value, the lower the transpiration; *kemis* is the soil emissivity used to compute land surface temperature. This parameter takes part in the net radiation calculation which determines the balance between incoming and outgoing energy at the surface; *kalbedo* weights the surface albedo, which is defined as the reflection coefficient for short wave radiation; *kcond* and *kcapa* take part in the thermal soil capacity and conductivity, both involved in the computation of the soil thermodynamics and *kz0* weights the roughness height, which determines the surface turbulent fluxes..
   1. Prior to the assimilation process, different scenarios were defined for the tests. A scenario makes reference to the experimental conditions. It includes the definition of the vegetation fraction (PFT), the type of observation to be assimilated, the observation sampling, the time sampling, and the atmospheric forcing file, subset of control parameters, assimilation window size and wished time of the year to start the assimilation. The different scenarios were calculated using the adjoint model for several typical summer conditions of the two Fluxnet sites selected. The dates presented in this paper (February 2 2003 for Kruger Park and August 26 1996 for Harvard Forest site) are representatives of sunny days in summer, with no perturbation coming from clouds and without rainfall events.
   2. In order to show the benefit of data assimilation in SECHIBA, we conducted several experiments using SECHIBA-YAO. The next section explains the scenarios for the different experiments performed in this work.

## 4.1 Variational sensitivity analysis

1. In order to show the accuracy of the distributed SECHIBA-YAO code, we present an analysis that allows to rank the eleven parameters according to their sensibility estimated by using the adjoint modeland compare the results to those obtained by using finite differences. We identify the most sensitive parameters to the estimation of land surface temperature by computing the gradients obtained with the adjoint model. This analysis corresponds to a first-order sensitivity estimate of the influence of the control parameters on the land surface temperature. In order to do so, local sensitivities were computed, providing the slope of the calculated model output variations in the parameter space for a given set of values (Saltelli et al, 2008). This method is really local and the information provided is related to a definite point in the parameter space. The values of the 11 parameters concerned in the analysis are presented in Table 1, they represent the initial values where the experiments have been conducted. Although humcste is related to vegetation type, in this work only value for PFT 0 (5 m-1) and PFT 11 (2 m-1) are considered.
2. The use of the SECHIBA-YAO software provided in this paper allows to test other locations using the dedicated data related to these locations.
   1. The sensitivity analysis was performed for a subset of inner parameters related to the energy and water physical processes on bare soil (PFT 0) and grassland (PFT 11), in order to quantify the role of the vegetation on the land surface temperature parameters’ sensitivity. The work was made on a daily basis, in order to observe the diurnal variations of sensitivities. At each half-hour time step, the model is restarted. At each time a gradient is computed in order to have the updated gradient value. Since no a priori values of the control parameters is known, as mentioned in section 2 , there is no background and the initial values of the parameters are those of table 1. For numerical purpose, in all scenarios the control parameters have been coded in order to have the same order of magnitude (equal to 1) during the minimization process.

Figure 4 compares, for August 26,1996 at Harvard Forest, the sensitivities computed for each control parameter with both finite differences and model gradients. Bare soil results are presented in Fig.4a. The grassland scenario is illustrated in Fig.4b. The efficiency of the adjoint calculation is first demonstrated in these plots, because the 11 desired parameters sensitivities are obtained in a single integration. The comparison between sensitivity analysis done using the adjoint and using finite differences shows a very good agreement between the two methods (the same results, not shown, were obtained with the Kruger Park site). The diurnal characteristics of the parameter sensitivities with a maximum around noon in line with the diurnal variation of solar radiation are clearly visible.

Table 2 presents, for Harvard Forest and Kruger Park, the 11 parameters ranked with respect to their influence. According to the four scenarios defined (two sites and two PFT), it can be seen that the hierarchy change with the vegetation, but remains the same for both sites. Parameter hierarchy revealed that the highest gradient values correspond to those that have the largest influence on the land surface temperature estimate. Clearly *kemis* is the most influential parameter in the calculation of land surface temperature, regardless of the climatology used and vegetation fraction. In addition, *mindrain* is the least influential parameter for all defined scenarios.

The parameters *kcapa, kcond, kzo* and *kalbedo* are the most influential in bare soil conditions, after *kemis*. In the presence of vegetation, several sensitivities change radically: *krveg* becomes the most important multiplicative factor after *kemis*; the factor *kalbedo* is less sensitive compared to its influence in the bare soil case and *mxeau* is more sensitive, given that less water is available when a fraction of vegetation is present. The other parameters show equivalent sensitivity values regardless the scenario. For *humcste* and *krveg*, sensitivities are equal to zero for bare soil, because these parameters affect surface temperature only in presence of vegetation.

Parameters with persistent positive sensitivity are: *rsolcste*, *krveg* and *humcste* . Parameters with persistent negative sensitivity are: *kz0*, *kalbedo* and *emis.* The sign of the gradients reflects the positive or negative feedback on the surface temperature of the processes involved. For example, the parameters involved in the evapotranspiration processes present negative sensitivities because a reduction (respectively an increase) of the evapotranspiration will lead to an increase (respectively a decrease) of the land surface temperature, when the soil water content is sufficient.

Transpiration processes influence directly the land surface temperature in presence of vegetation and is the dominant process in the studied sites. Therefore *krveg* has a higher sensitivity than *kcond*, *kcapa* and *kalbedo. .* For bare soil, on the contrary, the dominant processes are those related to the soil thermodynamics, explaining why *kcapa, kcond* and *kemis* are the most sensitive parameters*.* In general, sensitivities are higher in bare soil conditions for the control parameters, except for *mindrain* and *mxeau*.

The next section presents the different assimilation experiments that can be performed using the SECHIBA-YAO software.

## 4.3 Twin experiments

Twin experiments are synthetic tests checking the robustness of the variational assimilation method. The model is run with a set of parameters or initial conditions ***Ptrue*** in order to produce pseudo observations of land surface temperature ***TobsPtrue***. Then ***Ptrue*** is randomly noised to obtain ***Pnoise***. Assimilations of land surface temperature ***TobsPtrue*** were then performed in the model forced with ***Pnoise*** during several days (most of the time, one week), leading to a new set of optimized parameters denoted ***Passim***. Three different assimilation experiments were performed. These experiments are available in the distributed version of SECHIBA-YAO.

*Experiment Definition*

The parameters considered in the twin experiments are taken from the sensitivity analysis results. Each experiment was perturbed with a uniform distribution random noise reaching 50% of the parameter nominal value. We ran 500 assimilations in each experiment by randomly perturbing the initial conditions ***Pnoise*** (Table 1). This permitted us to obtain the relative errors of the control parameters and the relative values of the root mean square error (RMSE) of the model flux, based on their value before and after the assimilation process. The fluxes considered are net radiation (*Rn*), soil heat flux (*Q*), sensible (*H*) and latent heat (*LE*) fluxes between the atmosphere and the biosphere, land surface temperature (*LST*) and the soil water reservoir contents (*Water Stress*).

Scenarios for all the assimilation experiments are presented in Table 3. All parameters are controlled at the same time. The duration of each assimilation experiment is one week and the time increment is 30 minutes.

In Experiment 1 the five most sensitive parameters are controlled in bare soil conditions, during one week in Kruger Park and Harvard Forest sites.

In Experiment 2 the sixth most sensitive parameters are controlled in conditions of grassland (PFT 11) in Kruger Park and Harvard Forest site during a week.

With these two experiments, we are able to assess the effect of the vegetation fraction on the assimilation system. In addition, taking only the most sensitive parameters in the control set permitted to increase the assimilation performances, given that the more the observed variable is sensitive to a parameter, the easier the minimization process finds its optimal value, and consequently reducing the estimation error.

In Experiment 3, all parameters, except *mindrain*, are controlled (since *mindrain* has no impact in the land surface temperature estimation), during a week in both FLUXNET sites.

Comparing Experiment 3 with Experiments 1 and 2 allows us to study the impact of taking a larger control parameter set in the assimilation process. In addition, we want to test if land surface temperature as observation, provides enough information to constrain all the model parameters and if we can hope to improve all model state variables.

The RMSE errors of the 500 assimilations for each experiment are presented in Table 4 (resp Table 5) for Experiments 1, 2 and 3 corresponding to Harvard Forest (resp Kruger Park).

In Experiment 1, the errors on the retrieved values for all the control parameters are of the order of 10-13. Regarding the land surface temperature, the mean RMSE ranges from 0.03, prior assimilation, to 3.10-14 after the assimilation process. Same behavior is observed for the different model fluxes. Experiment 2 yields similar results as in Experiment 1. The assimilation process allows to reduce the parameter errors.

Several realizations of the prior parameters set in Experiments 1 and 2 (20% of the cases) did not converge at all. These results depend on the *a priori* parameter set value. They may indicate that a local minimum was reached in the minimization process of the cost function. In the case of non-convergence in Experiment 1, when the initial perturbations of *kcapa* and *kcond* differ (in sign or magnitude), the assimilation process does not reduce the prior errors. The correlation among *kcapa* and *kcond*, affects the convergence in the parameter space. This highlights the difficulty to characterize these two parameters independently with only observations of land surface temperature.

Relative value of the RMSE, with respect to the synthetic measurements, for *LE, H, Q* and *Rn* in Experiment 3 prior to assimilation, are equal to 34%, 30% and 23% and 0.6%, respectively. After assimilation, the RMSE is significantly reduced for both sites. The same holds for the mean relative error of the control parameters.

Comparing the results from Experiments 1 and 2 to Experiment 3, a degradation in fluxes and parameter restitution can be observed. Effectively, we find higher errors in the fluxes and the final control parameters when increasing the size of the control parameter set. Best performances in the parameters restitution are always for the control of 5 parameters, except at the Harvard Forest site, when the same performances were obtained with 5 or 6 control parameters. When controlling *krveg* plus 5 parameters in Kruger Park, a degradation is observed. This can be explained by the nature of the latter site: a higher shortwave incident radiation involves a smaller stomata resistance, and a larger transpiration, amplifying the weight of this flux in the temperature estimate. *krveg* sensitivity depends on the site and the vegetation fraction.

# 5. SECHIBA-YAO distributed version

The distributed version of SECHIBA-YAO provides an opportunity for scientists to perform their own assimilation. SECHIBA-YAO allows the control of the eleven most influent internal parameters of SECHIBA or of the initial condition. In iaddition soil temperature or satellite brightness temperature can be used as observations.

The distributed version of SECHIBA-YAO is available in LOCEAN website (ADRESSE ICI), the user must download the software, save it in a local repertory and run the *makefile* in order to build a local executable. Documentation and several instruction files are available in order to guide the user towards their own implementation.

The instructions files given with the distributed version correspond to the twin experiments presented in this paper. Initial parameters like the assimilation time frame and the observed variable (soil temperature) cannot be changed in the distributed version. However the other initial parameters to build different scenarios can be changed easily through the instruction file (initial parameter values, PFT, observations files, forcing, initial date, etc).

# 6. Conclusion

In this study the adjoint of SECHIBA was implemented, using an adjoint semi-generator software denoted YAO. With SECHIBA-YAO, land surface temperature gradients with respect to each control parameter were computed, with the aim at carrying out a sensitivity analysis of the parameter influence on LST estimation.

The first contribution of this paper is the sensitivity analysis results. They show exactly which parameters of the model are the most sensitive and have to be controlled during the assimilation process. However, it is important to mention that sensitivity analysis depends on the region, the forcing, the PFT, the day and night cycle, among other factors. The correlation study highlighted the interactions between parameters, giving insights into the model conceptualization, evidencing important parts of the model dynamics. Once the parameter hierarchy was set, twin experiments were performed for different scenarios, aiming at testing the robustness of the assimilation scheme.

The second contribution of this work is that we showed the usefulness of the variational data assimilation of LST to improve SECHIBA parameter estimations. Land surface temperature assimilation has the potential of improving the parameters of LSM, by adjusting properly the control parameters and the initial conditions. In a forecasting approach, this can be valuable, given that simulation can be more reliable since they are fitted on actual measurements. The improvement in the model fluxes after the assimilation of LST was demonstrated. Twin experiments showed the power of variational data assimilation to improve model parameter estimation. For different scenarios and forcing sites, the different experiments were successfully accomplished, meaning that a reduction in the fluxes errors was obtained by introducing information given by the LST synthetic observations. In addition, the influence that the size of the control parameter set has in the assimilation performance was shown.

Adding extra parameters to the control set increases the complexity of the cost function. Taking into consideration the results of assimilation of land surface temperature when controlling all parameter set (Experiment 3), we can see that, after having made ​​several assimilation runs, land surface temperature does not provide enough information to constrain all parameter set, in order to improve the estimation of state variables in SECHIBA. In the case of controlling all parameters we cannot hope improving all model state variables unless we assimilate additional parameters.

Assimilation with the YAO approach permits the implementation of different assimilation scenarios in a very flexible way, when performing different twin experiments: the control parameters and the observed variables (once the adjoint code has been generated), the assimilation windows, the observation sampling, the time sampling and other different features can be changed easily.

A distributed version of SECHIBA-YAO code and several examples with different scenarios are available at a LOCEAN dedicated site. YAO can be downloaded upon request. Direct use of this software will allow performing other experiments using different physical conditions or even changing several equations of the model itself.

# Acknowledgements

This work used eddy covariance data acquired by the FLUXNET community and in particular by the following networks: AmeriFlux (U.S. Department of Energy, Biological and Environmental Research, Terrestrial Carbon Program and AfriFlux). Dr. P. Peylin, F. Chevalier and M. Crépon are acknowledged for fruitful discussions. We thank also F. Maignan for its continuous support in the use of ORCHIDEE model, and Dr. M. Berrondo, for the assistance in writing this article.

# 7. References

Aubinet, M., Vesala, T., Papale, D. Eddy. Covariance: A Practical Guide to Measurement and Data Analysis. Springer Atmospheric Sciences Editions, United States of America. 2012.

Baldocchi, D ., Falge, E ., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw, K. T., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K., Wofsy, S. FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. Bull. Am. Meteorol. Soc., 82, 2415–2434. 2001

Barrett, D., Renzullo, L. On the Efﬁcacy of Combining Thermal and Microwave Satellite Data as Observational Constraints for Root-Zone Soil Moisture Estimation. CSIRO Land and Water, 1109-1127, Canberra, Australia. 2009.

Bateni, S.M., Entekhabi, D., Jeng, D.S. Variational assimilation of land surface temperature and the estimation of surface energy balance components. Journal of Hydrology, 481,143–156. 2013.

Benavides, H. Assimilation variationnelle des donnees dans le modèle de surface continentale SECHIBA grâce au logiciel YAO. These de doctorat de l’Université Paris 6. 2014.

Castelli F., Entekhabi, D., Caporali, E. Estimation of surface heat flux and an index of soil moisture using adjoint-state surface energy balance. Water Resources Research, 35, 10, 3115-3125. 1999.

d’Orgeval, T., Polcher, J., and Li, L. Uncertainties in modelling future hydrological change over west africa. Climate Dynamics, 26, 93–108. 2006.

Ducoudré, N., Laval, K., and Perrier, A. SECHIBA, a new set of parametrizations of the hydrologic exchanges at the land/atmosphere interface within the LMD atmospheric general circulation model. J. Climate, 6, 248–273. 1993.

Evensen, G. The ensemble Kalman filter: Theoretical formulation and practical implementation. Ocean Dyn., 53, 343–367. 2003.

Friedlingstein P., Joel G., Field C. B., Fung I. Toward an allocation scheme for global terrestrial carbon models. Global Change Biology, 5, 755-770. 1999.

Ghent, D. , Kaduk, J. , Remedios, J. and Balzter, H. Data assimilation into land surface models: The implications for climate feedbacks. International Journal of Remote Sensing, 3, 617 — 632. 2011.

Giering, R., Kaminski., T. Recipes for Adjoint Code Construction. ACM Transactions on Mathematical Software, 24, 437–474. 1998.

Gilbert, J.C., LeMaréchal, C. Some numerical experiments with variable-storage quasi Newton algorithms, Maths. Program, 45, 407-435. 1989.

Harrison, D.E., Chiodi A.M, Vecchi, G.A.Effects of surface forcing on the seasonal ycle of the eastern equatorial Pacific. J. Mar. Res. 67, 701-729. 2009.

Huang, C., Li, X., Lu, L. Retrieving land surface temperature profile by assimilating MODIS LST products with ensemble Kalman filter. Cold and Arid Regions Environmental and Engineering Research Institute, CAS, Lanzhou, China. 2003.

Ide, K., Courtier, P., Ghil, M. et Lorenc, A. Unified Notation for Data Assimilation : Operational, Sequential and Variational. Special Issue J. Meteorological Society Japan, 75, 181–189. 1997.

Kuppel, S., Peylin, P., Chevallier, F., Bacour, C., Maignan, F. and Richardson, A. Constraining a global ecosystem model with multi-site eddy-covariance data. Biogeosciences, 9, 3757–3776. 2012.

Lahoz, W; Khattatov, B. Data Assimilation Making Sense of Observations. Springer Editions. 2010.

Le Dimet, F.-X., Talagrand, O. Variational Algorithms for Analysis and Assimilation of Meteorological Observations: Theoretical Aspects. Dynamic Meteorology and Oceanography 38. 1986.

Nardi, L., Sorror, C., Badran, F., and Thiria, S. YAO: A Software for Variational Data Assimilation Using Numerical Models. Computational Science and its Applications - ICCSA 2009. International Conference, 5593, 2, 621-636. 2009.

Pipunic, R. C., Walker, J. P., and Western, A. Assimilation of remotely sensed data for improved latent and sensible heat flux prediction: A comparative synthetic study. 19th International Congress on Modelling and Simulation, Perth, Australia. 2008.

Reichle, R., Walker, J., Koster, R., Houser, P. Extended versus Ensemble Kalman Filtering for Land Data Assimilation. Journal of Hydrometeorology, 3, 728-740. 2001.

Reichle, R., Kumar, S., Mahanama, S., Koster, R. D., and Liu, Q. Assimilation of satellite-derived skin temperature observations into land surface models. Journal of Hydrometeorology, 11, 1103-1122. 2010.

Ridler, M., Sandholt, I., Butts, M., Lerer, S., Mougin, E., Timouk, F., Kergoat, L., Madsen, H. Calibrating a soil–vegetation–atmosphere transfer model with remote sensing estimates of surface temperature and soil surface moisture in a semi-arid environment. Journal of Hydrology 436–437, 1–12. 2012.

Robert, C, Blayo, E., Verron, J. Comparison of reduced-order, sequential and variational data assimilation methods in the tropical Pacific Ocean. Ocean Dynamics 56, 5-6 (2006) 624-633. 2007

Saltelli, A. Sensitivity Analysis. John Wiley & Sons Edition. United Stated of America. 2008

Sitch, S., Smith, B., Prentice, I.C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J.O., Levis, S., Lucht, W., Sykes, M.T., Thonicke, K., Venevsky, S. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. Glob. Change Biol. 9, 161 –185. 2003.

Table 1. SECHIBA Parameters studied in this work. There are 6 inner parameters, involved in the model estimations and 5 multiplying factors that are imposed to specific fluxes

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Description** | **Prior Value** | **Unit** |
| **Inner Parameters** | | | |
| humcste | Water stress | {5, 2} | m-1 |
| rsolcste | Evaporation resistance | 33000 | S/m2 |
| mindrain | Diffusion between reservoirs | 0,001 | - |
| dpucste | Total depth of soil water pool | 2 | m |
| mxeau | Maximum water content | 150 | Kg/m3 |
| **Multiplying Factors** | | | |
| kemis | Surface Emmisivity | 1 | - |
| kcapa | Soil Capacity | 1 | - |
| kcond | Soil Conductivity | 1 | - |
| krveg | Vegetation Resistant | 1 | - |
| kz0 | Roughness height | 1 | - |
| kalbedo | Surface albedo | 1 | - |

Table 2. Parameter hierarchy according to each site and vegetation fraction.

|  |  |  |  |
| --- | --- | --- | --- |
| **Site** | **Bare Soil (PFT 0)** |  | **Grassland (PFT 11)** |
| Harvard Forest | *kemis, kcond, kcapa, kz0, kalbedo, dpucste, rsolcste, mxeau* *mindrain, krveg humcste,* |  | *kemis, krveg, kcond, kcapa, kz0, mxeau, humcste, kalbedo, dpucste, rsolcste* *mindrain* |
| Kruger Park | *kemis, kcond, kcapa, kz0, kalbedo, dpucste, rsolcste, mxeau* *mindrain, krveg humcste,* |  | *kemis, krveg, kcond, kcapa, kz0, mxeau, humcste, kalbedo, dpucste,*  *rsolcste* *mindrain* |

Table 3. Scenarios for each of the 3 twin Experiments

|  |  |  |  |
| --- | --- | --- | --- |
| **Conditions** | **Experiment 1** | **Experiment 2** | **Experiment 3** |
| **Assimilation period** | 2 February 2003, 1 week (Kruger Park)  8 August 1996, 1 week (Harvard Forest) | 2 February 2003, 1 week (Kruger Park)  8 August 1996, 1 week (Harvard Forest) | 2 February 2003, 1 week (Kruger Park)  8 August 1996, 1 week (Harvard Forest) |
| **Number of assimilations** | 500 | 500 | 500 |
| **Control Parameters** | *kemis, kcond, kcapa, kz0, kalbedo* | *kemis, krveg, kcond, kcapa, kz0, kalbedo* | All parameters, except *mindrain* |
| **Observations** | Land surface temperature | Land surface temperature | Land surface temperature |
| **Observation sampling** | 30 minutes | 30 minutes | 30 minutes |
| **Forcing** | Kruger Park and Harvard Forest | Kruger Park and Harvard Forest | Kruger Park and Harvard Forest |
| **Vegetation type** | PFT 0 (Bare Soil) | PFT 11 (Grassland) | PFT 11 (Grassland) |

Table 4. RMSE of the Model fluxes and Relative errors of the parameters before and after the assimilation process for Experiment 1 to 3, on FLUXNET Harvard Forest

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Harvard Forest** | | | **Experiment 1** | | **Experiment 2** | | **Experiment 3** | |
|  | **Flux** | | **Prior** | **Final** | **Prior** | **Final** | **Prior** | **Final** |
| RMSE | Rn (W/m2) | | 2.23.10-1 | 1.27.10-19 | 2.31.10-1 | 1.37.10-7 | 6.33.10-3 | 4.90.10-5 |
| Q(W/m2) | | 1.32.10-1 | 3.28.10-12 | 1.33.10-1 | 2.86.10-6 | 2.31.10-1 | 4.18.10-2 |
| H(W/m2) | | 2.6.10-1 | 8.47.10-12 | 2.95.10-1 | 1.02.10-7 | 3.41.10-1 | 4.99.10-2 |
| LE(W/m2) | | 3.04.10-1 | 1.84.10-17 | 3.81.10-1 | 9.44.10-7 | 3.04.10-1 | 1.84.10-2 |
| Water stress (%) | | 1.29.10-3 | 6.03.10-19 | 1.18.10-5 | 3.16.10-8 | 6.34.10-3 | 4.90.10-5 |
| Temperature (K) | | 1.67.10-2 | 5.51.10-15 | 1.35.10-2 | 7.66.10-6 | 5.12.10-2 | 1.01.10-4 |
|  | |  | | | | | | | |
| Relative  Error | **Parameters** | | **Prior** | **Final** | **Prior** | **Final** | **Prior** | **Final** |
| kemis | | 2.55.10-1 | 5.71.10-6 | 2.42.10-1 | 5.96.10-7 | 2.63.10-1 | 2.1.10-3 |
| krveg | | - | - | 2.21.10-1 | 8.31.10-6 | 2.54.10-1 | 1.79.10-2 |
| kcond | | 2.4110-1 | 5.58.10-7 | 2.7.10-1 | 5.96.10-6 | 2.51.10-1 | 3.30.10-2 |
| kcapa | | 2.54.10-1 | 5.57.10-8 | 2.69.10-1 | 5.85.10-6 | 2.57.10-1 | 2.61.10-2 |
| kz0 | | 2.44.10-1 | 1.27.10-7 | 2.58.10-1 | 7.84.10-7 | 2.57.10-1 | 2.8.10-3 |
| kalbedo | | 2.44.10-1 | 1.99.10-6 | 2.39.10-1 | 2.08.10-6 | 2.47.10-1 | 2.37.10-3 |
| mxeau | | - | - | - | - | 2.58.10-1 | 7.34.10-2 |
| humcste | | - | - | - | - | 2.52.10-1 | 2.7.10-3 |
| dpucste | | - | - | - | - | 2.42.10-1 | 2.2.10-3 |
| rsolcste | | - | - | - | - | 2.54.10-1 | 2.36.10-3 |

Table 5. RMSE of model fluxes and Parameters Relative errors before and after the assimilation process for Experiment 1 to 3, on FLUXNET Kruger Park 02 February 2003

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Kruger Park** | | | **Experiment 1** | | **Experiment 2** | | **Experiment 3** | |
|  | **Flux** | | **Prior** | **Final** | **Prior** | **Final** | **Prior** | **Final** |
| RMSE | Rn (W/m2) | | 7.39.10-1 | 2.6.10-20 | 4.37.10-1 | 2.28.10-8 | 1.13.10-1 | 6.19.10-5 |
| Q(W/m2) | | 1.60.10-1 | 2.36.10-13 | 1.39.10-1 | 3.21.10-7 | 1.88.10-1 | 6.98.10-4 |
| H(W/m2) | | 4.55.10-1 | 8.47.10-13 | 4.73.10-1 | 1.80.10-7 | 4.46.10-1 | 1.46.10-3 |
| LE(W/m2) | | 1.51.10-1 | 8.84.10-17 | 1.24.10-1 | 8.86.10-7 | 3.61.10-1 | 1.01.10-2 |
| Water stress (%) | | 7.39.10-5 | 2.6.10-19 | 3.32.10-6 | 1.53.10-8 | 1.13.10-3 | 6.19.10-5 |
| Temperature (K) | | 3.41.10-2 | 1.62.10-15 | 2.75.10-2 | 1.51.10-7 | 1.54.10-2 | 3.4.10-6 |
|  | |  | | | | | | | |
| Relative  Error | **Parameters** | | **Prior** | **Final** | **Prior** | **Final** | **Prior** | **Final** |
| kemis | | 2.58.10-1 | 3.01.10-13 | 2.75.10-1 | 6.08.10-7 | 2.41.10-1 | 7.91.10-3 |
| krveg | | - | - | 2.81.10-1 | 2.76.10-8 | 2.29.10-1 | 4.91.10-3 |
| kcond | | 2.54.10-1 | 3.17.10-13 | 2.73.10-1 | 6.37.10-8 | 2.38.10-1 | 9.16.10-3 |
| kcapa | | 2.53.10-1 | 3.1.10-13 | 2.73.10-1 | 5.64.10-8 | 2.71.10-1 | 7.86.10-3 |
| kz0 | | 2.51.10-1 | 6.7.10-13 | 2.63.10-1 | 7.97.10-7 | 2.43.10-1 | 4.91.10-3 |
| kalbedo | | 2.59.10-1 | 5.2.10-13 | 2.63.10-1 | 2.31.10-6 | 2.53.10-1 | 3.47.10-2 |
| mxeau | | - | - | - | - | 2.46.10-1 | 6.16.10-3 |
| humcste | | - | - | - | - | 2.97.10-1 | 3.7.10-2 |
| dpucste | | - | - | - | - | 2.52.10-1 | 2.6.10-2 |
| rsolcste | | - | - | - | - | 2.41.10-1 | 1.26.10-2 |

Figure 1 (left) Example of a modular graph associated with four basic functions and five basic connections, three inputs points and three output points; (right) simplified description showing the acyclicity of the graph. Source: Nardi et al, 2009

|  |  |
| --- | --- |
|  |  |

Figure 2. (a) Example of a modular graph with five modules, assumed representative of the pointwise equations of a given model; (b) Partial view of the replication of the graph in space. Each elementary graph with five modules is associated with one grid point. Source: Nardi et al, 2009

|  |  |
| --- | --- |
|  |  |

Figure 3. Structure of a project in YAO. The software generates an executable program from input modules, hat and description files. The generated program reads an instruction file to perform assimilation experiments. Source: Nardi et al, 2009

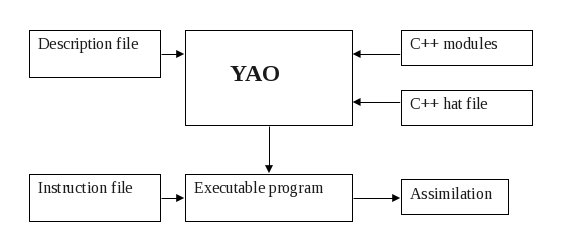


Figure 4 Comparisons for August 26,1996 at Harvard Forest, of the sensitivities obtained for each control parameter with both the finite differences and the model gradients computed with the adjoint model. Sensitivity analysis results for PFT 0 are in Fig.1 (a) and for PFT11 in Fig.1 (b). The sensitivities were computed on the surface temperature for Harvard Forest. Blue curves represent the LST derivative with respect to each parameter given by the adjoint each half hour over a day. Red curves represent the LST derivative computed with a finite difference discretization of the model.

|  |
| --- |
| (a) |
| (b) |

ANNEX

# SECHIBA-YAO

The version of SECHIBA implemented in YAO includes the two-layer hydrology of Choisnel (1977), mentioned in in Section 2. SECHIBA original code is implemented in a modular scheme, having a set of well-defined routines, independent in its processes and with a single entry point (a main routines handling the rest of the functionalities). The version of SECHIBA we used works in the vertical dimension only.

A set of prognostic variables is defined for each module and its assignation depends on the forcing conditions, physics phenomena, etc. SECHIBA can work coupled with the other components of ORCHIDEE (STOMATE and LPJ) or it can be used offline, as it was used in this work. Once SECHIBA is coded in YAO, it can be easily coupled with the other modules of ORCHIDEE.

In SECHIBA, the different routines were coded using Fortran language and can be run at any resolution and over any region of the globe. In the following, the version of SECHIBA implemented in YAO is denoted SECHIBA-YAO and the original version of the model, coded in Fortran, is denoted SECHIBA-Fortran.

ORCHIDEE uses MODIPSL and IOIPSL in its internal processes (see <http://forge.ipsl.jussieu.fr/igcmg/wiki/platform/documentation> for more information). Developed at IPSL, the first one is a set of scripts allowing the extraction of a given configuration from a computing machine and the compilation of the specific machine configuration components. MODIPSL is the tree that will host models and tools for configuration. IOIPSL helps to manage variables state history, variable normalization, file lecture, and among others.

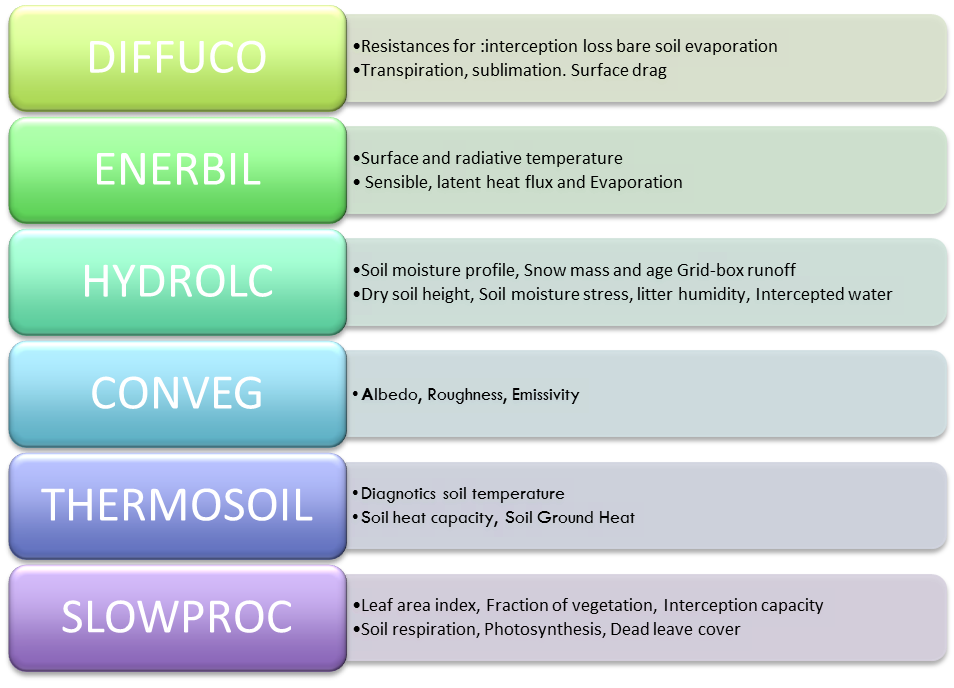


Figure A1 SECHIBA subroutines and its corresponding outputs. Source: Benavides, 2014.

The main routines in SECHIBA-Fortran are presented in Fig A1. These are also the routines considered in the YAO implementation of the model. First, DIFFUCO computes the diffusion and plant transpiration coefficients based on the atmospheric conditions, solar fluxes, dry soil height, soil moisture stress and fraction of vegetation. ENERBIL corresponds to the energy budget module. Surface energy fluxes related to the soil are computed, based on atmospheric conditions, radiative fluxes, resistances, surface type fractions and surface drag. HYDROLC is the hydrological budget module, taking as inputs the rainfall, snowfall, evaporation components, soil temperature profile and vegetation distribution. CONDVEG helps in the computation of the vegetation conditions. The thermodynamics of the model is computed in THERMOSOIL, based on a seven-layer soil profile. Finally, SLOWPROC computes the soil slow processes. When SECHIBA is decoupled from STOMATE, this module deals also with the LAI evolution.

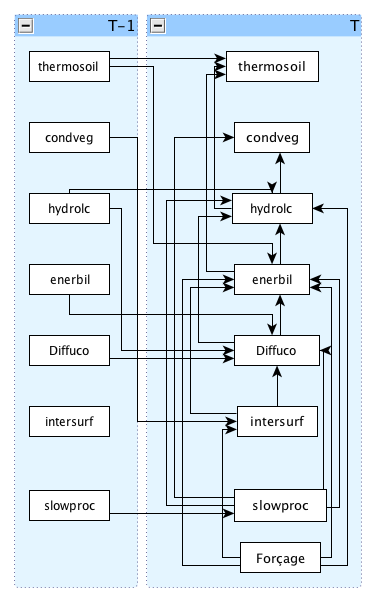


Figure A2 SECHIBA hyper graph, showing general model dynamics. Source: Benavides, 2014

The different SECHIBA components are interconnected as shown in Fig.A2. The output of the different modules serves as inputs for the next one, thus resulting in an interdependency among modules to be considered when modeling SECHIBA-YAO.