1. REPONSE TO BENOIT COUDERT

Dear Reviewer,

Thank you for your review and for the interest in our work. I make list of answers regarding all your comments and questions

SPECIFIC COMMENTS

Section 1:

P.2, L.31: could you precise what is a “specific deep land surface temperature” ?

The sentence has been changed in the revised version because it was wrong and related to another paper not referenced here. The new sentence is now: "When assimilating LST into the model, the authors proved that the assimilation of LST can improve the model simulated heat and water fluxes. "

P3, L.1: “or” should be replace with “of”

Modification taken into account

P3, L3: remove “available”

Modification taken into account

Section 2:

P4., L.34-36: The SECHIBA version used has a “two-layer soil profile” meanwhile in appendix A (P.28, L.8-9) a “seven-layer soil profile” is mentioned for the THERMOSOIL subroutine. Please bring some precisions or corrections.

A two-layer hydrology was used in this ORCHIDEE version. The seven layer discretization is for the resolution of the heat diffusion equation. We have changed the text in the paper to make it clearer

P.4.: L.1-12: could you precise why do you prefer the use of a brightness temperature in the interval [8-14] microns instead of the LST ? I certainly misunderstand the explanation.

The use of this variable follows my previous thesis work (Benavides, 2014) when observations coming from a thermal infrared radiometer were used as observations (SMOSREX). This interval correspond to the radiometer filter used for these measurements.

L.6, Eq.1: the Stefan Boltzmann constant [sigma] has been omitted in the first term of the equation.
L.6, Eq.1: is LW_down estimated or measured in situ ? In this case, could you precise the spectral band associated and if a band factor has been applied to take into account that only a fraction of

the radiation is measured in the spectral interval according to the Planck's law at the difference of the Stefan-Boltzmann law. Precisions are thus required regarding the use of the Svendsen conversion function (Eq.2).

> We don't understand your remark: in equation 1, we wrote the total radiation emitted by a soil surface and integrated on all the long wave spectra. The SB constant don't appear on the left side of the equation. In our case, LW downward is measured by a large band radiometer and this is why we can use the Svendsen's formula to estimate LST. The manuscript has been revised to clarify the notations and the confusions between LST and TB.

Table 3, P.18: "LST" is mentioned as observation but is it: LST, radiance or brightness temperature in the [8-14] microns interval ? You should also indicate that it is a synthetic observation.

I can assimilate LST or TB computed from a radiometer measurements. In my distributed version only LST observations are included. In the full SECHIBA-YAO version both measurements can be chosen.

P.4, L24: could you precise what is the type of the C3 crop for both sites and also give some details on the phenology or state of the plant development. As an example, LAI and canopy height could be added in Table 3 for PFT12.

Vegetation in ORCHIDEE is characterized by using Plant Functional Type system of classification. Although PFT system describes to types of cultures (C3 and C4crop) it does not distinguish varieties of crops and only one crop type is currently active

Section 3:

P.6, L.3, Eq.5: the cost function "f" should be replace with "J" in relation to Eq.4

> Modification taken into account

P.6, L.7: I suppose that "y" should be replaced with "J". I do not understand the reference to equation 2 which is the expression of the brightness temperature

> Reference to equation 2 misplaced. Modification taken into account

P.7, L.32: this empty line should be suppressed.

> Modification taken into account

P.8, L18-19: the sentence is unclear, please correct the syntax.

> The phrase will be replaced by: "When studying the subroutines, their complexity was reduced by breaking the different steps into simpler elements."

P.8, L.32: "the second approach was used" I certainly miss something but you have not presented several approaches in this subsection.

Misplaced reference: this sentence will be erased

Section 4: P.9, L.16: "The other parameters are multiplicative factors". Why don't you consider directly the parameters themselves: surface emissivity instead of kemis, albedo instead of kalbedo, etc. ? Is it only due to a technical (or numerical) reason ?

The idea is to have all parameters with the same value (all equal to 1) , in order to have directly the magnitude of the assimilation quality, and with the idea of having the possibility of comparing them

P.9, L.23: instead of "optimal value", you certainly mean "initial value" ?

What I meant is that prior to assimilation and to any perturbation, model parameters are always equal to 1

P.10, L.5-6 and Table 1 (P.16): the initial value of mxeau (maximum water content) parameter is very low (150kg/m3). Why this choice ?

This is the initial value generally used in sechiba before spinup.

What types of soil are considered? It is important to mention somewhere the soil description (classification or texture).

Yes , you are true, the soil texture has been added in the text .

A low mxeau value corresponds to dry or stressed surface conditions and will consequently increase the LST and overestimate it compared to in situ measurements. This remark is confirmed by the LE times series of figures 5&6 (see comments below) with quasi null absolute values. Is it done to increase the parameter sensitivity to LST in order to improve the results ?

Yes , we agree, and this is the case in our experiments , we took dry conditions to be close to the initial value prescribed in Sechiba, but we could have chosen another value. This is at this stage only synthetic observations and twin experiments. The next step is the assimilation of actual observations which will be our future work.

P.10, L.26-29: in order to facilitate the interpretation of the results of Figure 4 and Table 2, you should precise earlier how the parameter sensitivity hierarchy is defined with both methodologies (finite differences and model gradients), i.e. based on the slope of the gradients.

I didn't want to give much details on this because I think is out of the scope of the work: However I give a reference to my thesis (Benavides, 2014), where I give much details regarding this remark. However I clarified this point in the final manuscript

P.11, L.12-18: you should homogenize your notations throughout the text, tables and figures ("true" = observation, "noise" = first guest or perturbed, "assim"= after assimilation) in order to clarify.

Modification taken into account

P.12, section 4.4 “Results” and Tables 4 and 5: could you explain how a RMSD on LST reaching 5K is compatible with RMSD on surface fluxes lower than 2.5 W/m2 for experiment 1? The same could be addressed for experiment 2 although RMSD on LST is lower and RMSD on LE higher (but even though relatively low in absolute value). Figures 5&6: times series of LE for bare soil and although for C3 crop have very low absolute values (less than 5W/m2). It is related to the low mxeau value (see previous comment) ? Are the synthetic observations times series realistic compared to real observations ? You should give more information on these points in order to argue your choices and to comment the physical behavior of the model. From a physical point of view, I am surprised by the fact that times series are similar for figures 5 (bare soil) and 6 (C3 crop). During the simulation period of 7 days, LST increases by about 10K meanwhile H flux decrease and LE flux stays quasi null how is it possible ? Times series of meteorological forcing and a description of the vegetation development should be helpful for the analysis.

The experiments have been done in dry soil conditions , close to the initial value prescribed in Sechiba. We remind that we present here twin experiments, to present the tools developed and their potentialities. The dry soil conditions explain why there is not much difference between bare soil and C3crop with very low evapotranspiration rates. During this period, the ground heat flux increases and heat the soil, explaining the increase of the Surface temperature.

Section 5: P.13, L.1: “LST” should be replaced with “synthetic LST”.

done

2. REPONSE TO RIHAB MECHRI

Dear Reviewer,

Thank you for your review and for the interest in our work. I make list of answers regarding all your comments and questions

MINOR CORRECTIONS

• **Abstract**

1. The sentence corresponding to page 1, lines 16 to 18 is too long and should be shortened or divided in two sentences.

Remark taken into account. The phrase will be replaced from the manuscript to the following sentence: SECHIBA-YAO allows the control of the eleven most influent internal parameters and the initial conditions of the soil water content. This control is based on the assimilation of land surface temperature observations (in situ or from remote sensing as brightness temperature.

• Section 2 : Models and Data

1. Page 3, line 16 change “22th” to “22nd”.

Modification done

2. Page 4, line 10 the unit is not clear for the spectral band “um”.

It is the spectral band in micrometers (μm). Modification taken into account

• Section 3 : The Methodology

–Subsection 3.1: Variational Assimilation

1. Page 6, line 3 : you should replace the “f” at the end of equation (5) by “J”

Modification done

2. In the page 6, lines 6 and 7 you explained that y is described by equation 2. I can’t see the relationship between equation 2 describing the empirical formulation of the brightness temperature and the surface radiation and the description of the observation term “y”. Are you making reference to the equation described in page 5 at line 23 ?

Eq (2) makes reference to the calculation of brithness temperature in SECHIBA based on the empirical formulation of Svendsen et al., 1990. This variable can be later used as observation if remote sensing observations are used. The reference to equation 2 is misplaced, it will be remove from the manuscript

–Subsection 3.4: Development of SECHIBA-YAO

1. Page 8, line 11 : change "ANNEX A" to "Appendix A"

Modification done

• **Data assimilation experiments**

–Subsection 4.3: Experiment Definition

1. Page 11, line 27 : change as follows : "sensible (H) and latent (LE) heat

done

–Subsection 4.4: Experiment Definition

1. Page 12, line 25 : correct 'retrieved' to 'retrieve'.

done

QUESTIONS AND COMMENTS

Regarding the questions, I make a point by point answer to all your different comments.

1. In the variational assimilation can you please specify what do you exactly mean by observations and first guess : what are you exactly assimilating Pnoise (referred as 'first guess' and 'perturbed' in figures 5 and 6 (a and b)) or Ptrue (referred as 'observations' and 'initial value' in figures 5 and 6 (a and b)) ?

Since we are performing twin experiments, an initial set of parameters (Ptrue) is used to produce synthetic observations. The idea is to perturbate Ptrue (to obtain Pnoise,meaning my first guess). The idea is to used the synthetic observations produced with Ptrue in order to go from Pnoise to Ptrue by the assimilation process.

(a) In the case you are assimilation observations then how could you perform your validation using the same observations?

Since the assimilation process may give control parameters not exactly the same as the parameters wanted, its final values will affect the final model state, thus a comparison between observations and final temperature values can be useful

(b) In the case you are assimilation your Pnoise then can you explain more how did you perturbed the 'Truth' using your uniform random noise (precise the respective variation ranges of the different assimilated variables so that we can see how much 50% of the nominal value is consistent) ?

The perturbation was a random noise produce by the computer, limited up to 50% the true parameter value, equal to one, so the perturbed value is constrained between [0.5 , 1.5]

2. In the experiment 3 the gal was to show how could the number of variables included in the assimilation affects the performances of the method. In this case Experiment 3 must have the same conditions than Experiment 2 except the number of assimilated variables. Surprisingly you have

changed the assimilation period starting the 8th of August 1996 rather than the 3rd of March. My questions are the following :

(a) Why did you change the starting date of the assimilation?

I wanted to test different conditions in the assimilation capabilities. The scope of this work was only to demonstrate the potential of the assimilation tolos developed. For further information and tests, readers can consult my PhD manuscript (Benavides, 2014).

(b) How could you know that the decrease in the performances is only related to the number of parameters knowing that you have taken a different assimilation period and knowing the fact that the sensitivity of parameters toward LST is - as you have already mentioned- dependent on the seasons, period of the day etc. ?

The decrease in the performance is related to the complexity of the cost function to minimize: the greater the parameters the more complex will be and a decreased in the assimilation can be expected, regardless the season, period of the day, etc. In my PhD report (Benavides, 2014) I performed other experiments corroborating this statement.

3. REPONSE TO ABDELAZIZ KALLEL

Dear Abdelaziz,

Thank you for your review and for the interest in our work. I make list of answers regarding all your comments and questions

You said in section 3.1 that you use the Gradient algorithm but you do not explain what kind of algorithm it is exactly: is it "Levenberg-Marquardt algorithm" ?

> Regarding the gradient algorithm, a minimiser called M1QN3 is used within YAO. It use q quasi-Newton technique (the L-BFGS method of J. Nocedal) with a dynamically updated scalar or diagonal preconditioner.

You do not explain well how to estimate the actual control parameter values given the a priori. Indeed, the relationship prior value/actual value determines the covariance matrix B in Eq. (4)

> The Eq (4) is the most general form of the variational assimilation. I only give an introduction to this formula, but the estimation for the actual control parameter values are out of the scope of this work.

In your experiments you do not add noise to observation so in this case R is 0 and Eq.(4) is not well defined (division by 0). For that I suggest to add noise to observation an study the robustness of the developed approach as a function of the noise level.

> The equation 4 is just the general form. In YAO R is by default the identity matrix so users can modify its value when necessary

MODIFICATIONS TO THE MANUSCRIPT

Page 2: It is well known that both approaches provide the same solution at the end of the assimilation period, for perfect and linear models. - -> It is well known that both approaches provide the same solution at the end of the assimilation period, for Gaussian variables, and perfect and linear models.

> Modification taken into account

Page 5, Line 23: index i is forgotten in epsilon

> Modification taken into account

Page 8, line 32: you said "the second approach was used". I do not understand what is it "the second approach".

It refers to the type of coding of the modules in the modular graph. Since no detail is given before regarding this pointm the phrase will be erased from the manuscript

Page 9, line 11: you said "the initial model". Same problem, I do not understand.

It refers to the reference model, before parameter perturbation

Page 9, line 25: "the parameter prior values were retrieved successfully." In general, we estimate the actual values and not the prior. The prior is what we know initially before observation.

Exactly, but since is a twin experiment our prior is the target value we want to achieve. The phrase will be changed by: **"the assimilation was successfully achieved."**

Page 12, line 24: more difficult it is to find local minima that correspond to the initial control parameters values - -> more difficult it is to find global minima that correspond to the initial control parameters values

Modification taken into account

Page 12, line 25: It is difficult to retrieved parameters - -> It is difficult to retrieve parameters

Modification taken into account

Page 12, line 26: the assimilation of this variable in order to optimize these parameters is not optimal - -> the assimilation of this variable in order to optimize these parameters is not efficient

Modification taken into account

Variational Assimilation of Land Surface Temperature within the ORCHIDEE Land Surface Model Version 1.2.6

H. S. Benavides Pinjosovsky[1 2 3], S. Thiria[1], C. Ottlé[2], J. Brajard[1], F. Bradran[1] and P. Maugis[2]
[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}, Orsay, France

Correspondence to: H. S. Benavides Pinjosovsky (spinjosovsky@gmail.com) and S. Thiria (sylvie.thiria@locean-ipsl.upmc.fr)

Abstract. The SECHIBA module of the ORCHIDEE land surface model 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 was to deliver the adjoint model of SECHIBA (SECHIBA-YAO) obtained with YAO to provide an opportunity for scientists and end users to perform their own assimilation. SECHIBA-YAO allows the control of the eleven most influent internal parameters or initial conditions of the soil water content, by observing the land surface temperature or remote sensing data as brightness temperature.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. In addition, a distributed version is available when only the land surface temperature is observed.

Commentaire [BPHS1]: Modify Mme. Mechri

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 a 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 objective 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, given the model and observations uncertainties. Two basic methodologies can be used for that purpose. 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 Gaussian variables, and perfect and linear models. But both approaches become very different when the processes under study are highly nonlinear. 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.

Commentaire [BPHS2]: Modify M. Kallel

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 observations 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 the 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 errors. 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 When assimilating LST into the model, the authors proved that the assimilation of LST can improve the model simulated heat and water fluxes. As a conclusion, it can be seen that the assimilation of LST can improve the model simulated flows.

Commentaire [BPHS3]: Modify M. coudert

In the present study, we focused on the SECHIBA module (Ducoudré et al. 1993), part of the ORCHIDEE Land Surface Model, dedicated to the resolution of the surface energy and water budgets. Our objective was to test the ability of 4DVAR to estimate a set of its inner parameters as well as initial conditions of surface soil water content by observing the brightness temperature or the soil temperature. A dedicated software (denoted SECHIBA-YAO) was developed 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 to control the most influent internal parameters of SECHIBA by assimilating land surface temperature observations. At a given location and for specific soil and climate conditions, twin experiments of assimilation with remote sensing data can be executed. 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 assimilation tools are introduced in Section 5.

Commentaire [BPHS4]: Modify M. coudert

Commentaire [BPHS5]: Modify M. coudert

This 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.

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 biophysical processes. It is part of the IPSL earth system model (LMDZ, Hourdin et al., 2006), and is composed of 3 modules: SECHIBA, STOMATE and LPJ. The version used to this work correspond to the version 1.2.6, released the 22nd April 2010. 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 found 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 considered.

Commentaire [BPHS6]: Modify Mme. Mechri

In SECHIBA, 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 (R_n), 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 $h_{tot} = 2m$. 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: R_{lw} the incoming infrared radiation; R_{sw} the incoming solar radiation; P the total precipitation (rain and snow); T_a the air temperature; Q_a the air humidity; P_s the atmospheric pressure at the surface and U the wind speed.

Commentaire [BPHS7]: Modify M. coudert

In the full version of SECHIBA-YAO, observations of LST or brightness temperature can be used to constrain model inner parameter or initial conditions of the model variables. However, the simulated LST is hemispheric and does not account for solar configuration and viewing angle effects. In order to compute a thermal infrared brightness temperature from LST, and neglecting the directional effects, the total energy emitted by the surface (Rad) can be computed using the following expression :

$$Rad = k_{emis}\ \varepsilon\ LST^4 + (1 - \varepsilon\ k_{emis}) LW_{down} \qquad \text{(Eq 1)}$$

In this equation, ε is the surface emissivity, k_{emis} is the multiplicative factor for emissivity and LW_{down} is the longwave incident radiation that is an input forcing of SECHIBA. Svendsen et al. (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

Commentaire [BPHS8]: Modify Mme. Mechri

$$TB = \left(\frac{Rad - 7.84}{6.7975.10^{11}} \right)^{0.2} \qquad \text{(Eq 2)}$$

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 (CO_2), 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 agricultural C3 (PFT 12) and bare soil (PFT 1).

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 solution 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, $\mathbf{x}(t_0)$ is the initial state of the model and **k** represents the vector of the inner model parameters to be controlled, so $\mathbf{x}(t_i) = M_i(\mathbf{k}, \mathbf{x}(t_0))$, where $M_i(\mathbf{k}, \mathbf{x}(t_0))$ is represented by $M \circ M \circ \ldots \circ M(\mathbf{k},\mathbf{x}(t_0))$. The tangent linear model is noted $\mathbf{M}(t_i,t_{i+1})$, which is the Jacobean matrix of **M**, in $\mathbf{x}(t_i)$. The adjoint model $\mathbf{M}_i^T$ is the linear tangent transpose, defined as:

$$\mathbf{M}_i^T = \prod_{j=0}^{i-1} \mathbf{M}(t_j, t_{j+1})^T \qquad \text{Eq. (3)}$$

M is used to estimate variables, which are most often observed from an observation operator **H**, permitting to compare the observed values $\mathbf{y}^0$ with respect to the **y** calculated by the composition $\mathbf{H} \circ \mathbf{M}$, when they are available. The cost function *J* will be defined in terms of observations, so $\mathbf{H}_i$ allows us to estimate the variables $\mathbf{y}_i$, from the state vector $\mathbf{x}(\mathbf{t}_i)$. We suppose that $\mathrm{y}_i = \mathrm{H}_i(\mathrm{M}_i(\mathrm{x}_i, \mathrm{k})) + \varepsilon_i$ where ε_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:

Commentaire [BPHS9]: Modify M. Kallel

$$J(\mathbf{k}) = \frac{1}{2}\left(\mathbf{k}-\mathbf{k}^{b}\right)^{T}\mathbf{B}^{-1}\left(\mathbf{k}-\mathbf{k}^{b}\right)+\frac{1}{2}\sum_{i=0}^{t}\left(\mathbf{y}_{i}-\mathbf{y}_{i}^{0}\right)^{T}\mathbf{R}_{i}^{-1}\left(\mathbf{y}_{i}-\mathbf{y}_{i}^{0}\right) \quad \text{Eq. (4)}$$

The background vector is defined as $\mathbf{k}^{b}$, which is an *a priori* vector of the inner model parameters. The first part of the cost function represents the discrepancy to $\mathbf{k}^{b}$ 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 $\mathbf{k}^{b}$ and $\mathbf{R}_i$ is the covariance error matrix of $\mathbf{y}^{o}$ at time t_i. 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 present 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

$$\nabla_{k}J = \mathrm{B}^{-1}\left(\mathrm{k}-\mathrm{k}^{b}\right)+\sum_{i=1}^{t}\mathrm{M}_{i}^{T}(\mathrm{k})\nabla_{yi}J \quad \text{Eq (5)}$$

Commentaire [BPHS10]: Modify Mme. Mechri nad M. coudert

Where $\nabla_{k}J$ and $\nabla_{yi}J$ are the gradients of the cost function *J* with respect to **k** and $\mathbf{y}_i$ respectively.

The expression above allows us to compute $\nabla_{k}J$ by knowing $\nabla_{yi}J$, in the form of a matrix product of this term by the matrix $\mathbf{M}_{i}^{T}(\mathbf{x},\mathbf{k})$, corresponding to the transpose of the Jacobian Matrix. The development of calculation gives the expression of the gradient of **y**:

Commentaire [BPHS11]: Modify Mme. Mechri nad M. coudert

$$\nabla_{k}J = \mathbf{B}^{-1}\left(\mathbf{k}-\mathbf{k}^{b}\right)+\sum_{i=1}^{t}\mathbf{M}_{i}^{T}(\mathbf{k})H^{T}\left[R_{i}^{-1}\left(y_{i}-y_{0}\right)\right] \quad \text{Eq (6)}$$

The control parameters are adjusted several times until a stopping criterion is reached. The iterations of the gradient method allow us to approach the 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 scenario (incremental form) and is interfaced with the M1QN3 minimizer (Gilbert and Lemaréchal, 1989).

3.3 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 link a data input point located in the external context to 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 link 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 $F_p \rightarrow F_q$, then F_p should be computed before F_q) (see Fig.1)

Typically, a modular graph describes the equations governing the system of interest and each physical variable appearing in the governing equations are 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. Fig.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 data input points with a perturbation $\mathbf{dx}_i$ (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.

Following 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 back-propagating 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 to 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 gives access to a compiled version of SECHIBA-YAO and allows to perform some assimilation experiments related to the control of the ten most influent internal parameters of SECHIBA by observing the land surface temperature . YAO is a free software that gives the opportunity to modify the SECHIBA code provided in this paper.

3.4 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 Appendix 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 know 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, their complexity was reduced by

Commentaire [BPHS12]: Modify Mme. Mechri breaking the different steps into simpler elements. 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.

Commentaire [BPHS13]: Modify M. Coudert

-The original 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.

3.4.1 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.

Commentaire [BPHS14]: Modify M. Kallel, M. coudert

3.4.2 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 et al, 2012)). For SECHIBA-YAO, the derivative process was 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

In this section we present several experiments that have been realized using the SECHIBA-YAO.. They are related to the control of the eleven most influent internal parameters of SECHIBA by observing the land surface temperature.

The parameters are divided into two groups: inner parameters and multiplying factors (Table 1). The first group corresponds to physical parameters. The second group collects parameters weighting some physical processes of SECHIBA. In the initial model, they are all normalized to 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: $rsol_{cste}$ 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; hum_{cste}, mx_{eau} and min_{drain} 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; dpu_{cste} represents the soil depth in meters. The other parameters are multiplicative factors. We have k_{rveg} which is used in the calculation of the stomatal resistance, this variable limits the transpiration capacity of leaves, the greater its value, the lower the transpiration; k_{emis} controls the soil emissivity used to compute land surface temperature. This parameter takes part in the net radiation calculation which determines the energy balance between incoming and outgoing surface fluxes; k_{albedo} weights the surface albedo, which is defined as the reflection coefficient for short wave radiation; k_{cond} and k_{capa} take part in the thermal soil capacity and conductivity, both involved in the computation of the soil thermodynamics and k_{z0} weights the roughness height, which determines the surface turbulent fluxes. The control parameters are normalized from their prior value, so their optimal value is always equal to 1 and thus, only relative perturbations are considered. If the control parameter values posterior to the assimilation process are close to 1, it means that the assimilation was successfully achieved. Differences between the values retrieved and the prior values represent relative errors on the parameter estimation, posterior to assimilation.

Commentaire [BPHS15]: Modifi M. Kallel

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 functioning type (PFT), the type of observation to be assimilated, the observation sampling, the time sampling, and the atmospheric forcing file, the subset of control parameters, the assimilation window size and the 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 are representative of sunny days in summer or winter, with no perturbation coming from clouds and without rainfall events. 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

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 model and to 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 hum_{cste} is related to vegetation type, in this work only value for PFT 1 (5 m^{-1}) and PFT 12 (2 m^{-1}) are considered.

The sensitivity analysis was performed for a subset of inner parameters related to the energy and water physical processes on bare soil (PFT 1) and agricultural C3 crop (PFT 12), 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 step, a gradient is computed in order to have the updated gradient value. Since no prior 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. We recall that for numerical purpose, the control parameters have been normalized in order to have the same order of magnitude (i.e. 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.4(a). The agricultural C3 crop scenario is illustrated in Fig.4(b). The efficiency of the adjoint calculation is first demonstrated in these plots, because the 11 desired parameters sensitivities are obtained in a single integration. By using the same methodology, sensitivity curves were computed in the Fluxnet site Kruger Park, which are presented in Benavides (2014)

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). For more information, consult Benavides (2014), where the comparison between the two approaches is developed. The diurnal characteristics of the parameter sensitivities with a maximum around noon in phase 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 k_{emis} is the most influential parameter in the calculation of land surface temperature, regardless of the climatology used and vegetation fraction. In addition, min_{drain} is the least influential parameter for all scenarios.

The parameters k_{capa}, k_{cond}, k_{zo} and k_{albedo} are the most influential in bare soil conditions, after k_{emis}. In the presence of vegetation, several sensitivities change radically: k_{rveg} becomes the most important multiplicative factor after k_{emis}; the factor k_{albedo} is less sensitive compared to its influence in the bare soil case and mx_{eau} 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 hum_{cste} and k_{rveg}, 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: $rsol_{cste}$, k_{rveg} and hum_{cste} . Parameters with persistent negative sensitivity are: k_{z0}, k_{albedo} 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 k_{rveg} has a higher sensitivity than k_{cond}, k_{capa} and $k_{albedo.}$. For bare soil, on the contrary, the dominant processes are those related to the soil thermodynamics, explaining why k_{capa}, k_{cond} and k_{emis} are the most sensitive parameters.

In general, sensitivities are higher in bare soil conditions for the control parameters, except for min_{drain} and mx_{eau}. Since min_{drain} is not sensitive to the land surface temperature, this parameter is no longer controlled. Only the ten most influent parameters are used in the following sections.

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

4.2 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 ***Tobs***. Then ***Ptrue*** is randomly noised to obtain ***Pnoise***. Assimilations of land surface temperature ***Tobs*** 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.

4.3 Experiment Definition

The 10 most sensitive parameters are considered in the twin experiments (all parameters except min_{drain}). We present here the results obtained for one particular random perturbation of the parameters (the one provided in the distributed version, see Section 5). A statistic made with 500 different random realizations gave the same performances (Benavides, 2014). Each experiment was perturbed with a uniform distribution random noise reaching 50% of the parameter nominal value. We ran the assimilations in each experiment by randomly perturbing the initial conditions presented in 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 the land surface temperature (*LST*), sensible (*H*) and latent heat (*LE*)sensible (*H*) and latent (*LE*) heat.

Commentaire [BPHS16]: Modify Mme. Mechri

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 ΔT is 30 minutes. All experiments presented in this work use Harvard Forest as forcing. Same experiments are developed for Kruger Park site in Benavides (2014).

In Experiment 1 the five most sensitive parameters are controlled in bare soil conditions, according to the sensitivity analysis (Table 2), during one week in Harvard Forest site.

In Experiment 2 the five most sensitive parameters are controlled in conditions of agricultural C3 (PFT 12), according to the sensitivity analysis (Table 2), in 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 min_{drain}, are controlled (since min_{drain} has no impact in the land surface temperature estimation), during a week in Harvard Forest.

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 at the same time and if we can hope to improve all model state variables.

4.4 Results

The RMSE errors of the assimilations for experiments 1, 2 and 3 are presented in Table 4 (resp Table 5) corresponding to Harvard Forest site.

In Experiment 1, the errors on the retrieved values for all the control parameters are of the order of 10^{-8}. Regarding the land surface temperature, the RMSE ranges from 4.82 K prior assimilation, decreasing to $2.1.10^{-5}$ K 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 the reduction of the parameter errors (Fig.5 and Fig.6). Regarding the flux presented in both figures, it can be observed there are almost no difference between both series (for LE and H). This is caused by a dry soil with no precipitation during this week of the experiment, leading into a week evaporation and transpiration, inducing a week vegetation covering.

Commentaire [BPHS17]: Modify M. Coudert

Relative value of the RMSE, with respect to the synthetic measurements, for *LST*, *LE* and *H* in Experiment 3 prior to assimilation, are equal to 3.12 K, 34.1 W/m^2 and 30.4 W/m^2, respectively. After assimilation, the RMSE is 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, 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 (Experiment 3). Best performances in the parameters restitution are always for the control of 5 parameters. When we control the 10 most sensitive parameters, as in Experiment 3, degradation in the final value of the parameters is observed. This can be explained by the complexity of the model, the larger the control parameters set, the more difficult it is to find global minima that correspond to the initial control parameters values used to produce the synthetic observations It is difficult to retrieve parameters that are insensitive to LST, thus the assimilation of this variable in order to optimize these parameters is not efficient. .

Commentaire [BPHS18]: Modify M. Kallel

Commentaire [BPHS19]: Modify M. Kallel

Commentaire [BPHS20]: Modify M.Kallel

5. 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 synthetic LST estimation.

Commentaire [BPHS21]: Modification M.Coudert

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 time period (hour and day), among other factors. 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 LSM parameter calibration, by adjusting properly the control parameters. 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 the 10 most sensitive parameters (Experiment 3), we can see that, after having made several assimilation runs, land surface temperature does not provide enough information to constrain the 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 observations.

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 GitHub dedicated site. YAO can be downloaded upon request at https://skyros.locean-ipsl.upmc.fr/~yao/. Direct use of this software will allow performing other experiments using different physical conditions or even changing several equations of the model.

6. CODE AND DATA AVAILABILITY

The distributed version of SECHIBA-YAO provides an opportunity for scientists to perform their own assimilation. The distributed version allows the control of the 5 most influent internal parameters of SECHIBA, depending on the vegetation type. In addition, LST or satellite brightness temperature can be used as observations.

The distributed version of SECHIBA-YAO is available in a GitHub repository (https://github.com/brajard/sechibavar/archive/v1.0.zip), the user can download the software, save it in a local repertory and run the *makefile* in order to build a local executable. Documentation and two instruction files are available in order to guide the user towards their own implementation. Users can modify the forcing file, the initial date to the assimilation, the parameters value and their perturbation if needed. The assimilation frame (1 week), the step time (30 minutes), the observed variable (land surface temperature), the control parameters (only 5) and other initial parameters are imposed. If user wants to have access to a full modifiable version, YAO software has to be installed (https://skyros.locean-ipsl.upmc.fr/~yao/).

The instructions files given with the distributed version correspond to the twin experiments presented in this paper (Experiments 1 and 2). Initial parameters like the assimilation time frame and the observed variable (LST) cannot be changed in the distributed version. However the other initial parameters used to build different scenarios can be changed easily through the instruction file (initial parameter values, PFT, observations files, forcing, initial date, etc).

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 Dr. 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.

Hector Simon Benavides Pinjosovsky. Variarional data assimilation in the land surface model ORCHIDEE using YAO. Earth Sciences. Université Pierre et Marie Curie - Paris VI, 2014. English. <NNT : 2014PA066590>. <tel-01145923>. Available at http://www.theses.fr/2014PA066590

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.

F. Hourdin, I. Musat, S. Bony, P. Braconnot, F. Codron, J.-L. Dufresne, L. Fairhead, M.-A. Filiberti, P. Friedlingstein, J.-Y. Grandpeix, G. Krinner, P. LeVan, Z.-X. Li et F. Lott, 2006, The LMDZ4 general circulatiuon model : climate performance and sensitivity to parametrized physics with emphasis on tropical convection,

Climate Dynamics, 27 : 787-813

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.

Krinner, G., Viovy, N., Noblet-Ducoudre, N. de, Ogee, J., Polcher, J., Friedlingstein, P.,

Ciais, P., Sitch, S., and Prentice, I. C. A dynamic global vegetation model for studies of

the coupled atmosphere-biosphere system. Global Biogeochem. Cycles, 19. 2005.

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.

Parameter	Description	Prior Value	Unit
	Inner Parameters		
hum_{cste}	Water stress	{5, 2}	m^{-1}
$rsol_{cste}$	Evaporation resistance	33000	S/m^2
min_{drain}	Diffusion between reservoirs	0,001	-
dpu_{cste}	Total depth of soil water pool	2	m
mx_{eau}	Maximum water content	150	Kg/m^3
	Multiplying Factors		
k_{emis}	Surface Emissivity	1	-
k_{capa}	Soil Capacity	1	-
k_{cond}	Soil Conductivity	1	-
k_{rveg}	Vegetation Resistant	1	-
k_{z0}	Roughness height	1	-
k_{albedo}	Surface albedo	1	-

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

Site	Bare Soil (PFT 1)	Agricultural C3 crop (PFT 12)
Harvard Forest	k_{emis}, k_{cond}, k_{capa}, k_{z0}, k_{albedo}, dpu_{cste}, $rsol_{cste}$, mx_{eau} min_{drain}, k_{rveg} hum_{cste},	k_{emis}, k_{rveg}, k_{cond}, k_{capa}, k_{z0}, mx_{eau}, hum_{cste}, k_{albedo}, dpu_{cste}, $rsol_{cste}$ min_{drain}
Kruger Park	k_{emis}, k_{cond}, k_{capa}, k_{z0}, k_{albedo}, dpu_{cste}, $rsol_{cste}$, mx_{eau} min_{drain}, k_{rveg} hum_{cste},	k_{emis}, k_{rveg}, k_{cond}, k_{capa}, k_{z0}, mx_{eau}, hum_{cste}, k_{albedo}, dpu_{cste}, $rsol_{cste}$ min_{drain}

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

Conditions	Experiment 1	Experiment 2	Experiment 3
Assimilation period	()*+,s 1996, (Harvard Forest)	3 Mars 1996 1 week (Harvard Forest)	8 August 1996, 1 week (Harvard Forest)
Control Parameters	k_{emis}, k_{cond}, k_{capa}, k_{z0}, k_{albedo}	k_{emis}, k_{rveg}, k_{cond}, k_{capa}, k_{z0}	All parameters, except min_{drain}
Observations	Land surface temperature	Land surface temperature	Land surface temperature
Observation sampling	30 minutes	30 minutes	30 minutes
Vegetation type	PFT 1 (Bare Soil)	PFT 12 (Agricultural C3crop)	PFT 12 (Agricultural C3crop)

Table 3. Scenarios for each of the 3 twin experiments

	Experiment 1 (PFT 1)				Experiment 2 (PFT 12)			
	Relative error (%)		**RMSE**		**Relative error (%)**		**RMSE**	
	Prior	**Final**	**Prior**	**Final**	**Prior**	**Final**	**Prior**	**Final**
LST (K)	5.2	$3.1.10^{-10}$	4.82 K	$2.1.10^{-5}$ K	7.78	$1.35.10^{-6}$	1.61 K	1.10^{-10} K
LE(W/m^2)	5.10	$5.1.10^{-6}$	2.5 W/m^2	$6.6.10^{-4}$ W/m^2	13.56	$1.2.10^{-5}$	8.52 W/m^2	$1.2.10^{-6}$ W/m^2
H(W/m^2)	2.53	$1.59.10^{-8}$	2.03 W/m^2	1.1^{-12} W/m^2	39.23	$1.3.10^{-3}$	1.39 W/m^2	$1.2.10^{-10}$ W/m^2

(a)

	Relative error (%)			
	Experiment 1 (PFT 1)		**Experiment 2 (PFT 12)**	
	Prior	**Final**	**Prior**	**Final**
k_{emis}	14.69	0	20.92	$5.019.10^{-3}$
k_{z0}	28.18	0	48.42	$6.81.10^{-3}$
k_{cond}	44.99	0	38.8	$3.23.10^{-3}$
k_{capa}	48.98	0	11.48	$7.32.10^{-3}$
k_{rveg}	-	-	44.83	$1.69.10^{-3}$
k_{albedo}	38.25	$2.384.10^{-7}$	-	-

(b)

Table 4. Results for Experiments 1 (PFT 1) and 2 (PFT 12). RMSE of model fluxes (a) and Parameters Relative errors (b) before and after the assimilation process on FLUXNET Harvard Forest, 03 Mars 1996 during a week

	Experiment 3 (PFT 12)			
	Relative error (%)		**RMSE**	
	Prior	**Final**	**Prior**	**Final**
LST (K)	5.12	$1.1.10^{-3}$	3.12 K	$3.2.10^{-1}$ K
LE(W/m^2)	7.10	$5.2.10^{-2}$	34.1 W/m^2	3.1 W/m^2
H(W/m^2)	2.53	$2.39.10^{-2}$	30.4 W/m^2	2.1 W/m^2

(a)

	Relative error (%) (PFT 12)	
	Experiment 3	
	Prior	**Final**
k_{emis}	26.3	$2.1.10^{-1}$
k_{z0}	25.4	$1.79.10^{-1}$
k_{cond}	25.1	$3.30.10^{-1}$
k_{capa}	26.7	$2.61.10^{-1}$
k_{rveg}	27.5	$2.8.10^{-1}$
k_{albedo}	24.7	$2.37.10^{-1}$
mx_{eau}	25.8	$7.34.10^{-1}$
hum_{cste}	25.2	$2.7.10^{-1}$
dpu_{cste}	24.2	$2.2.10^{-1}$
$rsol_{cste}$	25.4	$2.36.10^{-1}$

(b)

Table 5. Results for Experiment 3 (PFT 12). RMSE of model fluxes (a) and Parameters Relative errors (b) before and after the assimilation process, on FLUXNET Harvard Forest, 08 August 1996 during a week

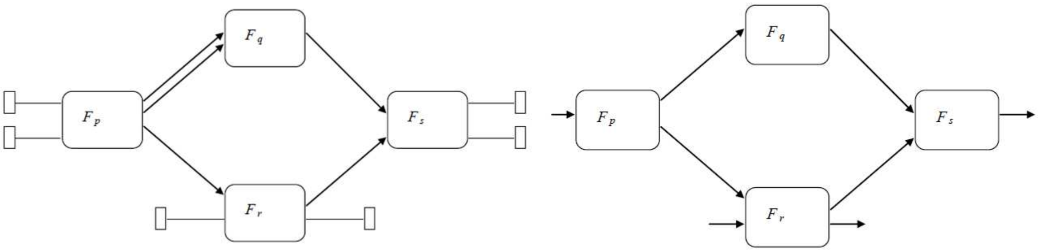


Figure 1 (left) Example of a modular graph associated with four basic functions and five basic connections, three inputs points and three output points; (right) simplified description showing the acyclicity of the graph. Source: Nardi et al, 2009

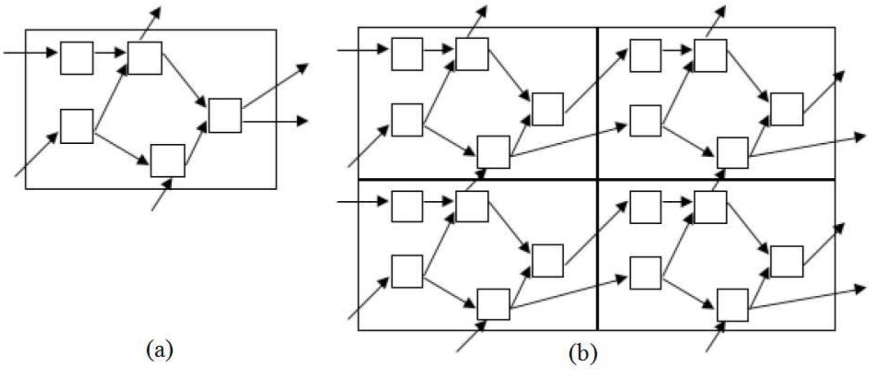


Figure 2. (a) Example of a modular graph with five modules, assumed representative of the pointwise equations of a given model; (b) Partial view of the replication of the graph in space. Each elementary graph with five modules is associated with one grid point. Source: Nardi et al, 2009

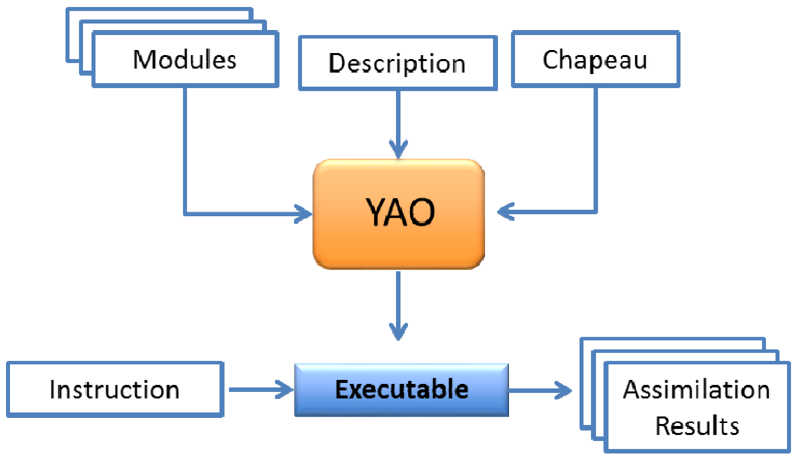


Figure 3. Structure of a project in YAO. The software generates an executable program from input modules, hat and description files. The generated program reads an instruction file to perform assimilation experiments.

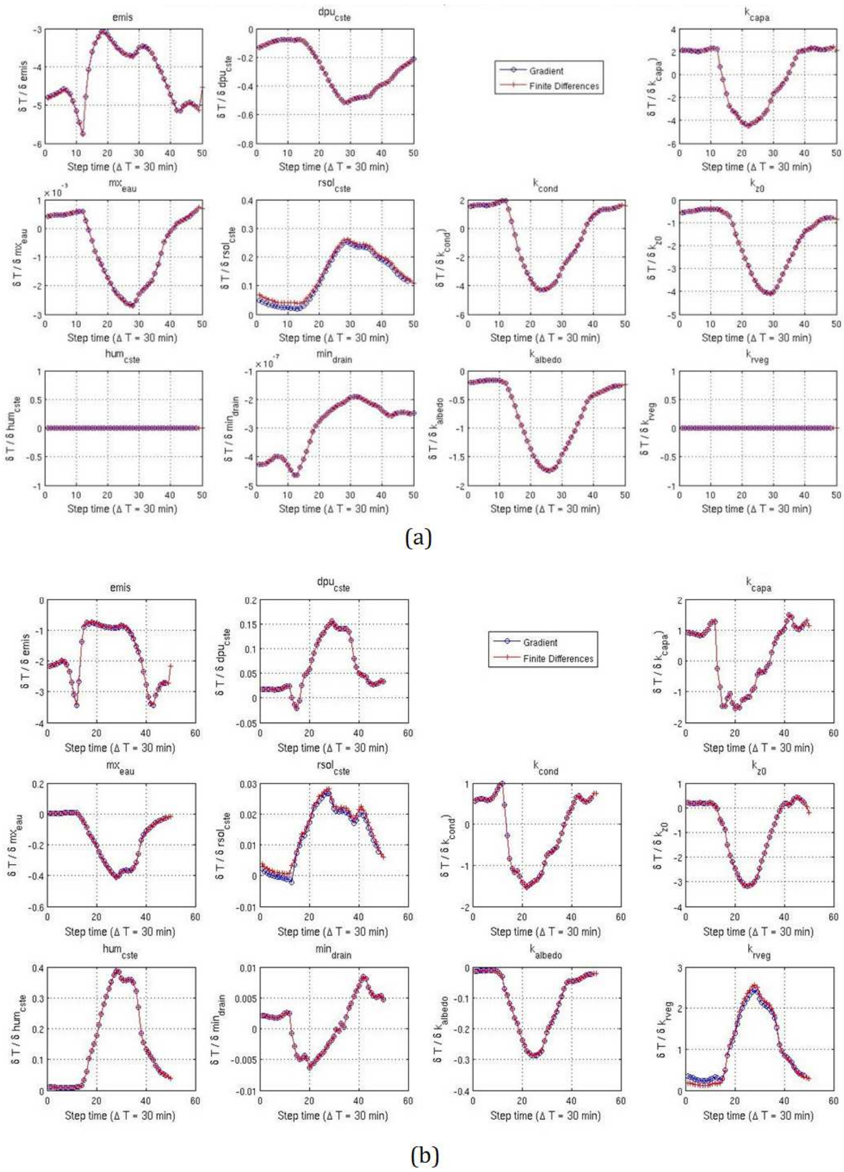


Figure 4. Comparisons for August 26,1996 at Harvard Forest, of the sensitivities obtained for each control parameter with both the finite differences and the model gradients computed with the adjoint model. Sensitivity analysis results for PFT 1 are in Fig.4 (a) and for PFT 12 in Fig.4(b). The sensitivities 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.

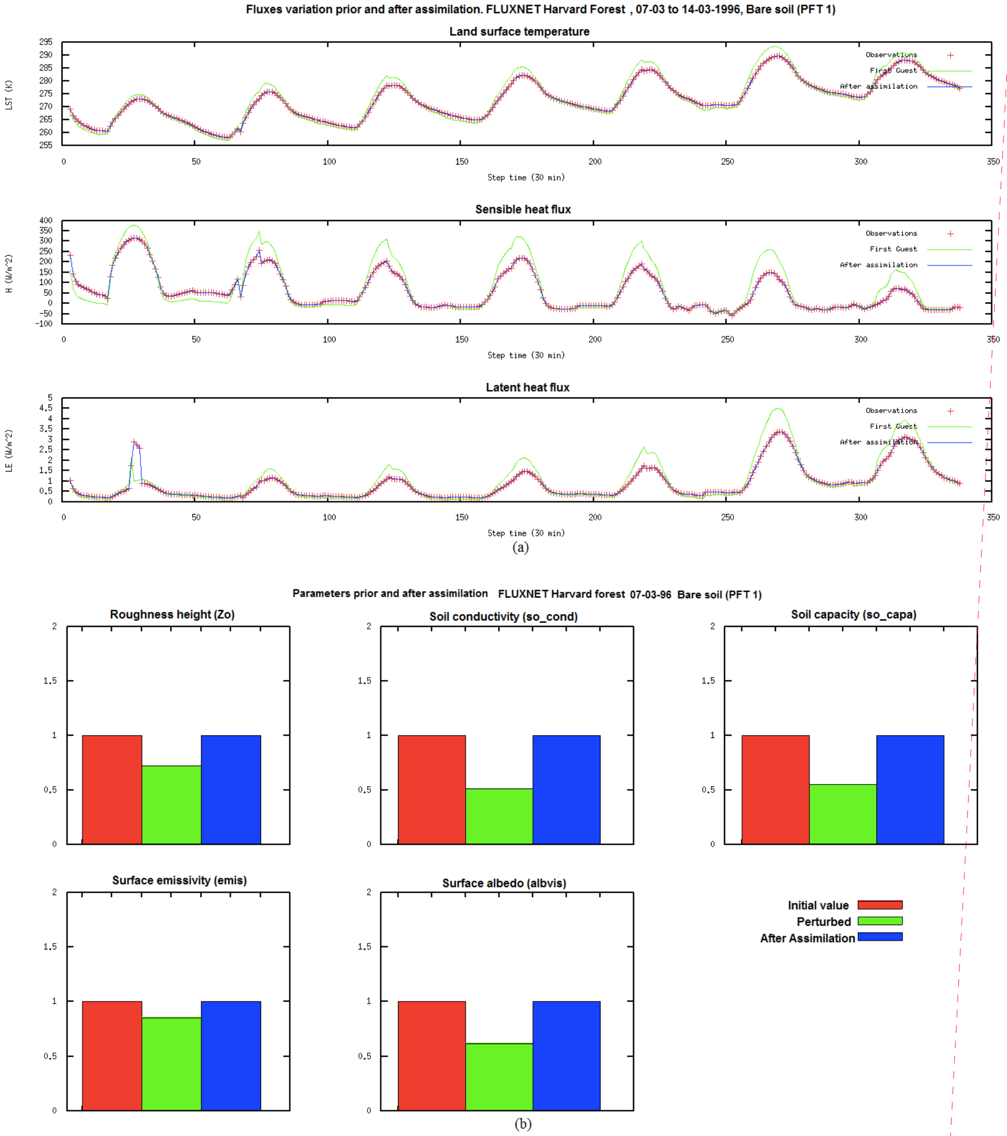
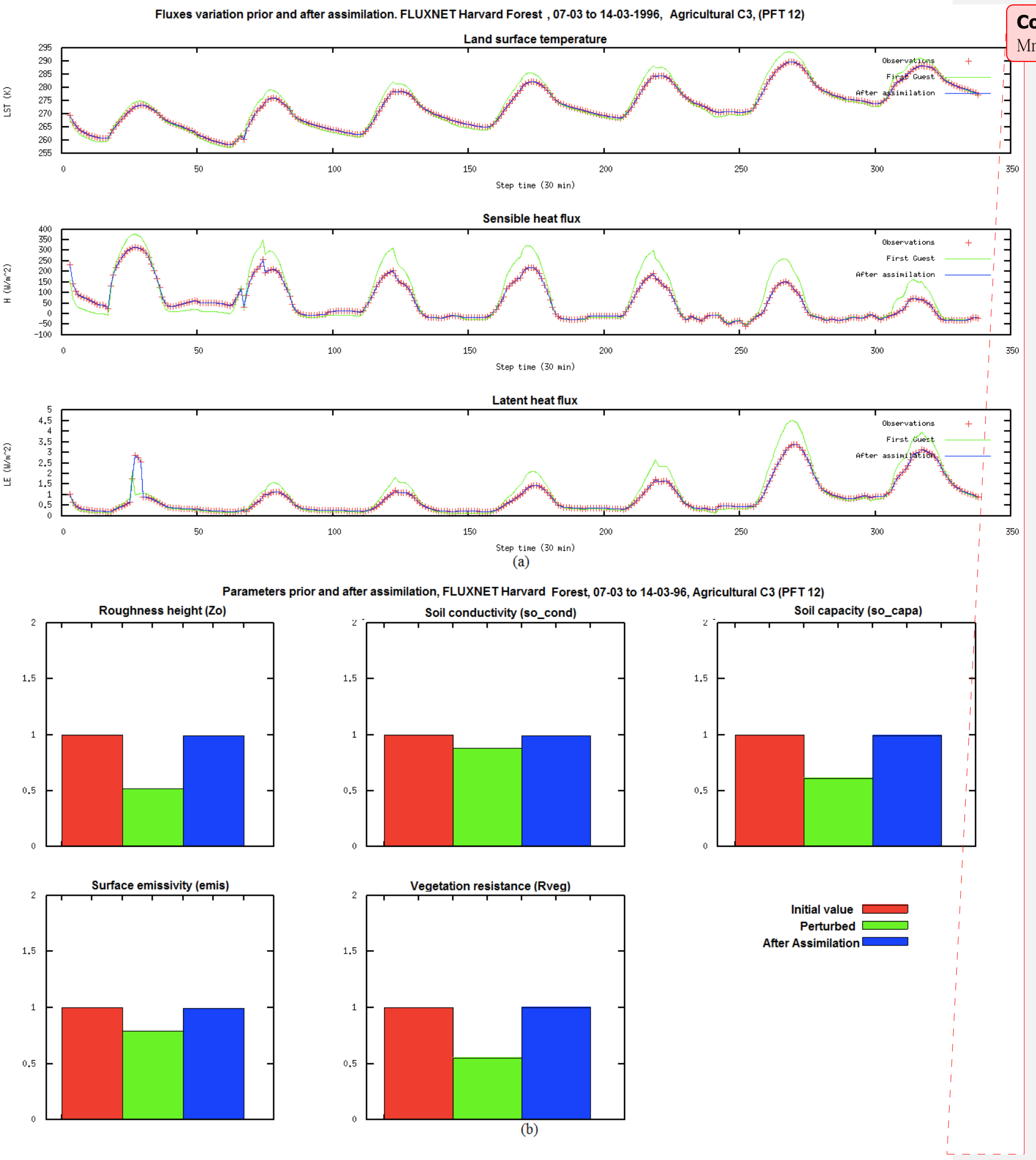


Figure 5. Comparison between variables and parameters prior and after assimilation, for experiment 1. LST, H and LE are compared in Fig. 5.(a) and parameters values in Fig.5(b). Parameters values after assimilation corresponds to values used to produce the synthetic observations and thus validating the twin experiment.

Commentaire [BPHS22]: Modify Mme. Mechri



Commentaire [BPHS23]: Modify Mme. Mechri

Figure 6. Comparison between variables and parameters prior and after assimilation, for experiment 2. LST, H and LE are compared in Fig. 6.(a) and parameters values in Fig.6.(b). Parameters values after assimilation corresponds to values used to produce the synthetic observations and thus validating the twin experiment.

APPENDIX A

SECHIBA-YAO

The version of SECHIBA implemented in YAO includes the two-layer hydrology of Choisnel (1977), mentioned 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).

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. It can be run only one point at a time. ?

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.

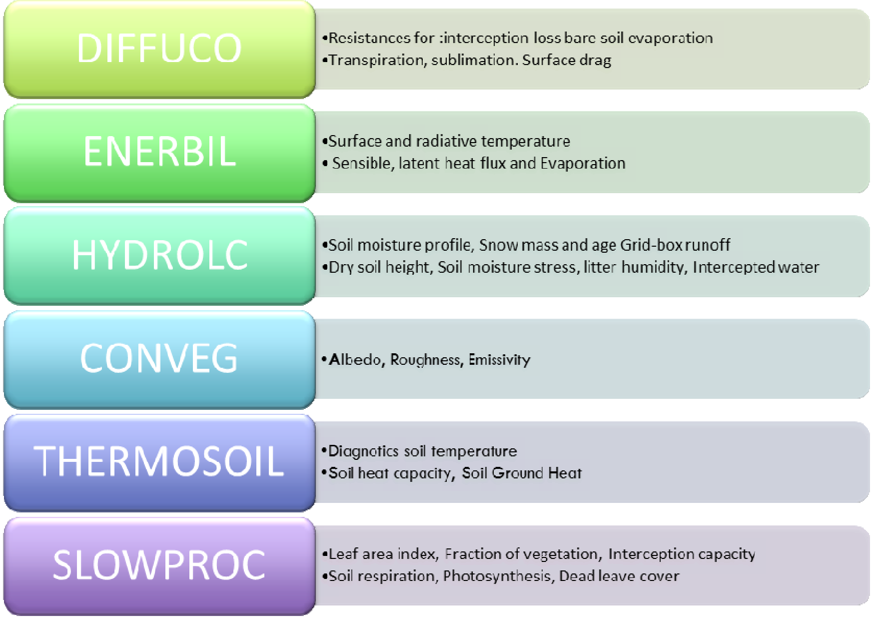


Figure A1 SECHIBA subroutines and its corresponding outputs. Source: Benavides, 2014.

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.

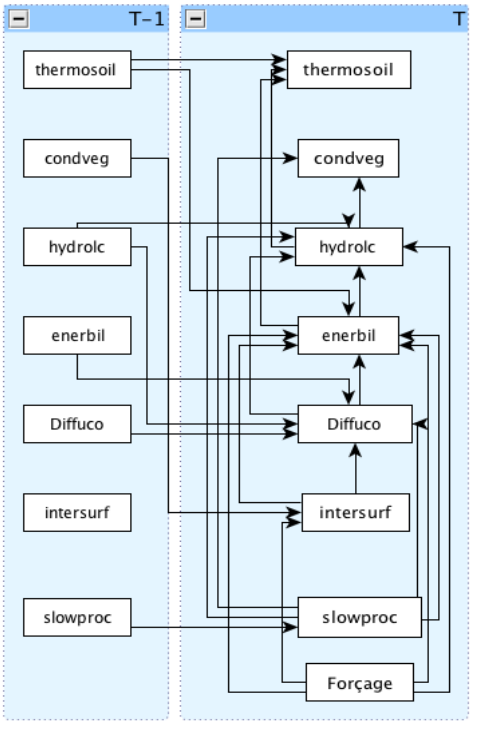


Figure A2 SECHIBA hyper graph, showing general model dynamics. Source: Benavides, 2014

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.