Coupled assimilation of in situ flux measurements and satellite fAPAR time series within the ORCHIDEE biosphere model: constraints and potentials

C. Bacour[1,2], P. Peylin[3], P. Rayner[2], F. Delage[2], M. Weiss[4], J. Demarty[5], F. Baret[4], F. Chevallier[2], P. Prunet[1]

[1]*NOVELTIS, Ramonville-Saint-Agne, France*
[2]*Laboratoire des Sciences du Climat et de l'Environnement (LSCE), Gif-sur-Yvette, France*
[3]*Laboratoire de Biogéochimie des milieux continentaux, Thivernal-Grignon, France*
[4]*Unité Environnement Méditerranéen et Modélisation des Agro-Hydrosystèmes (EMMAH), Avignon, France*
[5]*HydroSciences Montpellier, Université de Montpellier 2, Montpellier, France*

cedric.bacour@noveltis.fr

Abstract

This study evaluates to compatibility of eddy flux tower measurements with time series of fAPAR derived from satellite and *in situ* measurements, in assimilation into the ORCHIDEE biosphere model.

1. Introduction

Our understanding of the role of the Earth's biosphere in the carbon cycle remains subject to large uncertainties. The process-based models used to simulate the response of the terrestrial ecosystems to environmental changes at the global scale rely on generic hypothesis and parameterizations. Spatial and temporal variations of the ecosystem characteristics are still poorly represented whereas they may yield to divergent modeled responses to a same environmental constraint. Assimilation of *in situ* eddy flux data allow improving the model parameterization at the local scale, the measurements being representative of the exchange of mass and energy between ecosystem and atmosphere within only a few hundred meters around the flux towers. Because of their high spatial coverage and revisit frequency, assimilation of satellite-derived biophysical products into terrestrial ecosystem models is foreseen to also substantially improve to model trajectories and parameterizations. This however assumes that satellite products and model outputs are actually compatible.

Using the state of the art mechanistic vegetation model ORCHIDEE, we have investigated at the Laboratoire des Sciences du Climat et de l'Environnement (LSCE) the benefit of combining remotely sensed fAPAR (fraction of absorbed photosynthetically active radiation) and *in situ* flux measurements to improve the model simulations.

2. Materials and methods

2.1. The ORCHIDEE vegetation model

ORCHIDEE is based on the concept of plant functional types. It is run here at local scale to compute the energy, carbon and water balances on a half-hourly basis, depending on the meteorological forcing and the biome composition of studied sites. Besides, the phenology scheme is fully prognostic thus allowing the derivation of the time course of LAI (leaf area index) and fAPAR (Krinner et al., 2005).

2.2. Sites studied

We focus on three instrumented sites for which half-hourly eddy covariance measurements of net CO2 (NEE), sensitive heat (H), and latent heat (LE) fluxes, are available: Fontainebleau (temperate deciduous forest), Le Bray (evergreen needleleaf forest), and Puechabon (evergreen Mediterraneous forest). Complementarily, fAPAR estimates were derived from medium (MERIS) and high (SPOT) spatial resolution sensors using an artificial neural network approach, and/or from in site measurements. In order to palliate the scarcity of High spatial Resolution (HR) images due to cloud cover, the available observations were extrapolated in time with a double sigmoid model, and corrected from the seasonality of the MERIS data.

2.3. fAPAR products

Large differences appeared between the *a priori* ORCHIDEE fAPAR simulated values and the various products and with MODIS collection 5 products at 1km. They are of the same order of magnitude as the differences between the different satellite products and the compatibility between ORCHIDEE fAPAR and each of these products varies across sites, as seen in Figure 1. These differences are twofold: differences in extreme fAPAR levels and differences in temporal profile.

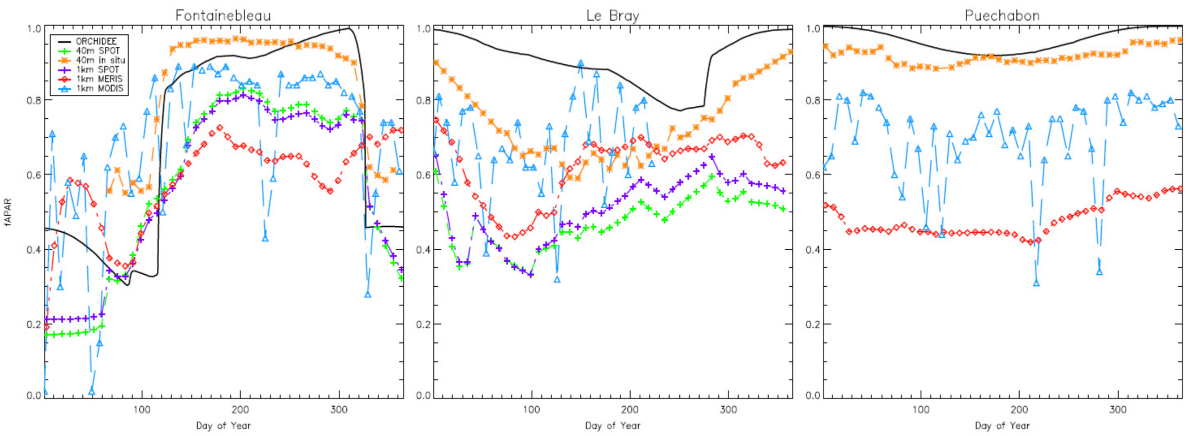

Figure 1 Comparison of the various fAPAR products, with MODIS collection 5 fAPAR products at 1km, and with prior ORCHIDEE simulations, for Fontainebleau, Le Bray, and Puechabon.

At any spatial resolution, the satellite products are systematically lower than the prior model simulations and the *in situ* measurements. The differences in spatial heterogeneity of the land surface may explain this feature for the 1km products (for Puechabon, MERIS fAPAR is up to the half of the *in situ* measurements and the modelled values). Even though some heterogeneity may occur at high spatial resolution, such discrepancy is rather a consequence of the fAPAR processing chain (estimation algorithm, saturation effect, atmospheric correction, etc.). At the opposite, the levels of the *in situ* fAPAR products are very close to those of the prior model simulations.

The differences also expresses with regard to seasonality. They are even more manifest for the deciduous forest site of Fontainebleau that presents the most pronounced seasonal cycle. The phasing of leaf onset and senescence is similar for ORCHIDEE and the *in situ* observations, and both show abrupt slope changes at these phenological stages. The satellite products however present smooth temporal variations the origin of which is twofold: fAPAR processing and spatial heterogeneity. Note that the MODIS collection 5 products presented here show a higher temporal variability as they were not temporally smoothed. ORCHIDEE represents the behavior of one idealized "plant". Therefore the model can not account for spatial heterogeneities within and between species that occur in a real landscape. Such heterogeneity leads to a temporal variability of leaf onset and senescence, giving rise to a smooth time course for fAPAR, as seen by the satellite products, MERIS in particular. The timing of the growing season also differs, the satellite derived fAPAR indicating a start of the growing season in advance as compared to ORCHIDEE prior simulation or *in situ* measurements. For Fontainebleau, 1km MERIS products present high values at the beginning and end of the year: they are due to faulty atmospheric correction and high oblique illumination angles beyond the domain of validity of the training dataset used to train the neural network estimation algorithm. For Puechabon, both simulations and fAPAR products present a bowl shape temporal variation with higher values in winter: for ORCHIDEE simulations, it is due to the dependency of fAPAR to the variation of the sun zenith angle as 1/cos; but it can also reflect leaf adaptation to dry summer conditions and a reducing of the photosynthetic activity.

2.4. Assimilation setup

A four dimensional variational data (4D-var) assimilation system has been developed in order to assimilate the various fAPAR products, synergistically with ecosystem flux measurements, and thus optimize some ORCHIDEE parameters related to carbon assimilation, respiration and phenology, processes. The approach relies on the minimization of a Bayesian misfit function *J(x)* that measures the mismatch between 1) seasonal cycle of actual observations *Y* and corresponding model outputs *M(x)*, and 2) *a priori* x_p and optimized parameter *x* values, weighted by the prior error covariance matrix on observations *R* and parameters *B* (Tarantola, 1987):

$$J(X) = (Y - M(X))^t R^{-1} (Y - M(X)) + (X - X_p)^t B^{-1} (X - X_p)$$

The determination of the optimal set of parameters that minimizes *J(x)* is performed with the L-BFGS-B algorithm (Zhu *et al.*, 1995), specifically dedicated to solve large nonlinear optimization problems subject to simple bounds on the variables. The gradient of the misfit function, required by the optimization algorithm, is computed by the tangent linear version of ORCHIDEE.

We analyze the compatibility of the various fAPAR products with ORCHIDEE's prognostic phenology and with the *in situ* flux measurements, by assimilating each type of data separately or together. The fluxes assimilated are net CO2 (NEE), sensitive heat (H), and latent heat (LE); for readability however, only the time series of NEE and fAPAR are presented in the following.

3. Assimilation results

For Fontainebleau, there is a relatively good agreement of the prior ORCHIDEE model with the observed flux data (phase and amplitude of the seasonal cycle), except for the sensible heat flux. The assimilation of the fluxes significantly improves the model-data fit and leads to a very good overall model performance for NEE (Fig.2). For Puechabon and Le Bray, the flux model simulations perform slightly worse than for the previous deciduous site, especially for the simulation of the hydrological stress on NEE during a dry season (for Le Bray this corresponds to the summer 2003 heat wave).

As a consequence of the large differences between the *a priori* ORCHIDEE fAPAR simulated values and the various products, the fluxes simulated after assimilation of these satellite fAPAR data strongly depart from the prior values: for most cases, the agreement between flux measurements and ORCHIDEE simulations has worsened. The *in situ* fAPAR measurements are in closer agreement with the prior model (both in terms of level and timing of the growing season).

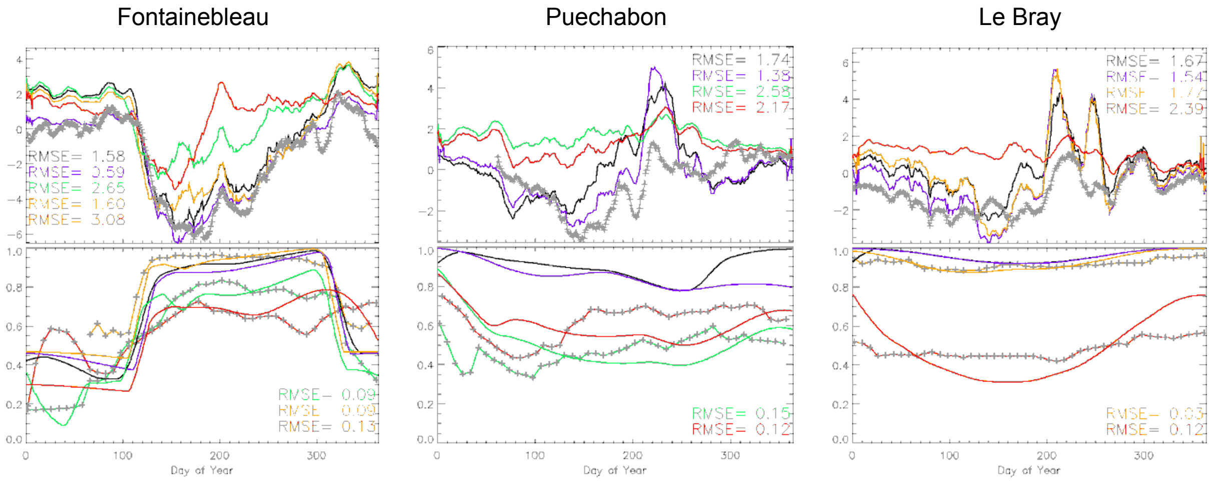

Figure 2: Comparison between NEE flux and fAPAR observation data and the corresponding ORCHIDEE model simulations prior to assimilation, and posterior to assimilation of only flux data, and only fAPAR data considering HR (pseudo-SPOT data corrected from the seasonality from MERIS MR observations, and in situ *measurements) and MR (MERIS products at 1km) time series, for Fontaintebleau, Puechabon and Le Bray. The RMSE gauges the quality of the fit.*

The results obtained when assimilating the two pieces of information together (flux measurements and fAPAR products) are a compromise between those obtained using them separately (data not shown). The optimizations are mostly driven by the phasing of the NEE seasonal cycle, the posterior simulations remaining close to the flux data. As compared to the case where only flux data are assimilated, there is only a light degradation of the posterior model-data fit, mostly in winter. Despite the strong discrepancy between fAPAR levels between the satellite products and the prior model simulation in summer, the reduction of the fAPAR level in the posterior model has almost no impact on the corresponding NEE values. This compromise indicates that there is a set of parameters that can accommodate for the large NEE seasonal cycle and a maximum fAPAR (or LAI) smaller than in the prior simulation.

4. Conclusion

The use of satellite fAPAR data (either HR or MR products) helped to constrain some critical parameters controlling the phenology of ORCHIDEE (maximum LAI, turnover of the leaves, time of leaf onset,…), some of them being poorly constrained by the use of flux data only. Differences between the parameters derived from the assimilation of HR, MR, or *in situ* fAPAR data only, were often physically and statistically significant. These findings indicate that, independent of the respective accuracy of each product, current differences between HR and MR data have a significant impact on model parameters and subsequently on modeled fluxes. HR fAPAR data potentially bring more local-scale cloud-free than MR data for a combined use of flux and fAPAR information, which is crucial considering the large spatial heterogeneity of the land cover associated to anthropized landscapes (in particular Europe and North America).

Overall, we saw substantial benefit in the use of HR data for ecosystems/sites where ORCHIDEE works relatively well in terms of simulated NEE and LE fluxes. However, in these cases the benefit of HR data is conditioned on the availability of high temporal frequency of the satellite products. For extra-tropical ecosystems with a pronounced growing season, any useful space observing system should be able to assess these temporal variations on at least a weekly basis. For evergreen ecosystems, the need for accurate mean level of fAPAR appeared as the first requirement for an effective use of the satellite products within ecosystem models. For deciduous or managed ecosystems, the temporal coverage of the product, especially during the leaf onset, leaf senescence, transition between wet and dry seasons, etc., become the most critical requirement.

5. References

Krinner G., Viovy N., de Noblet-Ducoudre N., Ogee J., Polcher J., Friedlingstein P., Ciais P., Sitch S., Prentice I.C. (2005). A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, Global Biogeochemical Cycles, 19(1), GB1015.

Tarantola A. (1987), Inverse problem theory: Methods for data fitting and parameter estimation. Elsevier, Amsterdam.

Zhu C., Byrd R.H., Lu P., Nocedal J. (1995), A Limited Memory Algorithm for Bound Constrained Optimization, (1995), SIAM Journal on Scientific and Statistical Computing , 16, 5, pp. 1190-1208.