Supplementary Information for "Assimilation of multiple datasets results in large differences in regional to global-scale NEE and GPP budgets simulated by a terrestrial biosphere model"

Bacour C., MacBean N., Chevallier F., Léonard S., Koffi E.N., Peylin P.

Supplementary Text S1: Data assimilation experiments: differences with the stepwise approach

Although the stepwise assimilation has been extended to ten years of atmospheric CO_2 data (compared to only three years in Peylin et al. (2016)), there are few differences in the experimental set-up compared to the DA experiments considered in the present study: *i*) the set of optimized parameters is not strictly identical: the stepwise study did not optimize the parameters controlling the maximum LAI value per PFT nor the root profile, which are included in this study, but instead did include two additional parameters, one controlling the albedo of vegetation and the other reducing the hydric limitation of photosynthesis, which are not considered here; *ii*) for optimization of the phenology using satellite NDVI data, the C4 grass PFT was calibrated in Peylin et al. (2016), which is not the case in this study (MacBean et al., (2015) found that phenology for semi-arid PFTs was not well captured by the model and further improvements to the phenology schemes for these PFTs are needed).; iii) the selection of the eddy-covariance sites is more selective in the present study (a few sites for which the model-data inconsistency was too important were discarded), which slightly reduces the number of site-years available for some PFTs; iv) finally, the *a priori* errors on model parameters (at the first and second steps) were set to 40% of the parameter variation range in Peylin et al. (2016), and were hence larger than what is prescribed in this study as a result of the consistency checks performed in Section 2.3.4.2.

Supplementary Text S2: Processing of atmospheric CO_2 data

In order to analyze the fit to the atmospheric CO_2 concentrations in terms of trend and seasonal cycle (magnitude and phase), the measured and modeled monthly time series are fitted using the CCGCRV package (ftp://ftp.cmdl.noaa.gov/user/thoning/ccgcrv/) following Thoning et al. (1989). It decomposes the time series into a first-order polynomial term (that represents the trend) and four harmonics, and then filters the residuals of that function in frequency space using a low-pass filter (cutoff frequency of 65 days). The seasonal cycle corresponds to the harmonics plus the filtered residuals. For a given time series, we calculate the magnitude of the seasonal cycle for each year as the difference between the maximum and minimum value, and the carbon uptake period (CUP) as the sum of the days when the values of the seasonal cycle extracted from the CO_2 concentration time series are negative (plant removing CO_2 from the atmosphere by convention) (Peylin et al., 2016). Examples of observed and simulated time series of atmospheric CO_2 concentrations at four sites are provided on Figure S1, as well as the corresponding trends derived by CCGCRV.

Supplementary Text S3: Consistency diagnostics on the errors

Desroziers et al. (2005) tests

Several attempts were performed to specify the errors on model parameters in order to approach this goal considering each data-stream independently. With an initial definition of the parameter error corresponding to 40% of their variation range, the diagnostics on the **R** matrix, show a strong overestimation for all data streams (ratios about 3 for NEE and LE, 2 for NDVI and 12 for atmospheric CO_2), while the diagnostics on **B** were more consistent with ratios slightly higher than 1 but for NDVI (2.5). These results led us to revise the definition of **B** by decreasing the error for all parameters such that it corresponds to about 20% of the variation range for phenological parameters, and 12% for the other parameters (a value close to what was prescribed in Kuppel et al. (2013).

Reduced chi-square

For all experiments but those involving atmospheric CO2 measurements, the values of the reduced chi-square (after optimization) over all data are below 1 (Table S1), which corroborates the overestimation of the model-data and parameter errors observed previously. For fluxes and satellite data, this overestimation of the model-data error was expected, and even desired, given that the covariances in R were neglected by construction (off-diagonal elements set to zero). For CO2, the large value of χ^2 expresses a strong underestimation of the observation error not highlighted by the consistency diagnostics. Indeed, when determining $\mathbf{R}_{\mathbf{CO2}}$, we purposely did not account for the structural error in ORCHIDEE that largely explains the strong bias between observed and simulated CO2 temporal profiles by about 1 $ppm.yr^{-1}$. This underestimation is even inflated in the joint assimilation experiments, even though the reduced chi-square over all data remains close to 1.

	Data - stream			
experiment	**F**	**VI**	**CO2**	**all data**
F	0.91			0.91
VI		0.78		0.78
CO2			8.57	8.57
F+VI	0.95	0.50		0.73
F+CO2	0.97		11.56	1.4
VI+CO2		0.74	11.72	1.18
F+VI+CO2	0.99	0.75	11.3	1.09
F+VI+CO2-2steps	0.96	0.67	7.88	0.97

Table S1: Values of the reduced chi-square determined after model calibration for the various assimilation

experiments, for each data-stream.

Supplementary Text S4: Optimisation performances

All optimizations but two (F and VI) reached a pre-defined maximum number of iterations (set to 35 for L-BFGS-B), therefore causing a hard stopping of the optimization (cf, Table S2, which also provides the values of the misfit functions for all assimilation experiments, relative to the background and to the observations). For the last iterations however, the variations of the misfit functions were low in all these cases, indicating that the final iterations were close to the minimum. The comparison between the observation and parameter terms of the posterior cost function shows how the total cost function is dominated by the weight of the model-data misfit.

The highest rate of change of the total cost function related to the observation term is obtained for the CO2 assimilation with a reduction of the misfit between model outputs and measurements by about 46. This is directly related to the correction of the large bias in the prior model with carbon pools at equilibrium relative to the prescribed prior error. Noticeably, the strong model improvement reached for CO2 comes with only a small variation in the model parameters as depicted by the posterior value of J_b. For the assimilation of the fluxes and satellite data alone (F and VI respectively), the model improvement is smaller, about 1.1, but shows a stronger departure of the parameters from their prior values compared to CO2 (Figure 3). The ratio of the norm of the gradient of the misfit function is also the highest for the CO2 experiment. On the opposite, it is slightly lower than one for VI which may indicate a possible issue of convergence towards the solution.

The two-step approach for the assimilation involving the three data-streams results in an enhanced agreement of the model with all data as compared to the one-step optimization. In parallel, the change in parameter values (departure from the background) is also higher for the two-step approach (Figure 3).

experiment	Number of iterations	Jo prior	Jo post	Jo prior/ Jo post (obs part)	Jo(F) post	Jo(CO_2) post	Jb post	Ratio norm grad J (prior/post)
F	34	75396	68305	1.10	68305		117.6	3.95
VI	29	65696	58517	1.12			37.9	0.94
CO2	35	1256783	27238	46.14		27238	7.8	759.5
F+VI	35	142118	108961	1.30	71353		79.3	0.97
F+CO2	35	1332190	109994	12.11	73232	36763	1.05	27.7
VI+CO2	35	1323494	92543	14.30		37257	1.3	132.3
F+VI+CO2	35	1398901	166797	8.39	74435	35918	1.6	168.7
F+VI+CO2-2steps	35	1398901	148206	9.43	72654	25002	44.6	-

Table S2: Characteristics of the various assimilation experiments: number of iterations, value of the cost functions related to the observation (Jo) and parameter terms (Jb) prior and posterior to the assimilation (as well as ratio of the posterior to prior values for Jo), ratio of norm of the gradient of the misfit functions (prior vs

posterior).

Supplementary Text S5: Analysis of the reduction of the model-data misfit

Mono-data stream assimilations

The increased consistency between model and flux data achieved after assimilation of F data is usually higher for NEE (median RMSD reduction of 10.4%, ranging from -69% to 38%) than for LE (0.3%; -42% / 28% range). This is largely explained by the higher number of optimized parameters related to the carbon cycle relative to the water cycle, and by the optimization of the multiplicative factor of the soil carbon pools that corrects the bias in the ecosystem respiration inherent to the model spin-up (Carvalhais et al., 2010; Kuppel et al. 2012). The strong model improvement for FAPAR in the VI assimilation (22.2% median; -32% / 36% range) follows a strong decrease of the simulated growing season length for deciduous PFTs in better accordance with the satellite observations, as discussed in MacBean et al. (2015). It mainly results from an earlier senescence for the several PFTs while the change of leaf onset depends on the type of vegetation. Both for the F and VI experiments, the reduction of the model-data misfit can be negative for some sites/pixels. This reflects how the assimilation may degrade the model performance at some sites/pixels by seeking for a common parameter set. This is not observed for atmospheric CO_2 data for which the optimized model is always closer to the observations than the prior model at all stations. Assimilating atmospheric CO_2 concentration measurements corrects the strong overestimation of the prior model (as also described in Peylin et al. (2016)), with a median RMSD reduction of 76% (ranging from 10% at HUN to 90% at SPO). This improvement corresponds to an increase of the net land carbon sink at the global scale in order to correct the strong mismatch between the observed trend and the *a priori* model. It is mainly realized by the optimization of the multiplicative factor of the soil carbon pools. As seen in Figure 2 from the detrended seasonal cycles of atmospheric CO_2 data (light red box), the changes in the modelled amplitude and phasing is smaller but still in better agreement with the observed data (median value of RMSD reduction of 14.4%; -21% / 55% range).

Multiple-data stream assimilations

The simultaneous assimilation of flux measurements and satellite NDVI data leads to enhanced model improvement as compared to when these data are assimilated alone: the median RMSD reductions are 10.8% for NEE (10.4% in the F case) and 36.7% for FAPAR/NDVI (22.2% in the VI case). In the simultaneous assimilations involving atmospheric CO_2 data, the most of the model improvement is attributed to CO2 while the benefit relative to fluxes and FAPAR/NDVI is weak: for NEE, the median RMSD reductions are only of 2.5% and 2.6% in the F+CO2 and F+VI+CO2 cases (as compared to 10% in the F case); for FAPAR, the median values are 1.2% and 1.4% for the VI+CO2 and F+VI+CO2 experiments

(22% in the VI case).

The 2-steps assimilation F+VI+CO2 results in a higher model improvement regarding both NEE and FAPAR (respectively 5.5% and 11.2%) than the one-step approach.

Regarding the raw atmospheric CO_2 data, the median improvements are 76.1% for CO2, 76.3% for F+CO2, 73.6% for VI+CO2, 72.9% for F+VI+CO2 and only 25.6% for F+VI+CO2-2steps.

More pronounced differences between experiments are obtained for the de-trended CO_2 time series: while the median RMSD reduction is of 14% in the CO2 experiment, it is decreased to 7.8% in F+CO2, 8.4% in VI+CO2, and 10.6% in F+VI+CO2; at the opposite the RMSD reduction is increased to 15.4% in F+VI+CO2-2steps.

<u>Supplementary Text S6: Global budget and uncertainty reduction</u>

For NEE, the global scale budget is about -2.4 $GtC.yr^{-1}$ for all experiments using atmospheric CO_2 as a constraint: the lower value of -2.28 $GtC.yr^{-1}$ is found for F+CO2; the higher values of -2.49$GtC.yr^{-1}$ and -2.48 $GtC.yr^{-1}$ are obtained for CO2 / F+VI+CO2-2steps.

In the northern and southern hemispheres, the CO2 assimilation results in the largest C sinks (-1.65 / -0.04 $GtC.yr^{-1}$ for NH/SH) while the 2step assimilation induces the lowest one (-0.41 / 0.003 $GtC.yr^{-1}$); the opposite result is obtained in the southern hemisphere with lower (-0.79 $GtC.yr^{-1}$) / higher (-2.06 $GtC.yr^{-1}$) budgets found for CO2 / F+VI+CO2-2steps.

The reduction of the global scale GPP budget is respectively of -19.61 $GtC.yr^{-1}$ and -17.91 $GtC.yr^{-1}$ for the F and VI experiments, which correspond to the largest corrections obtained among the various assimilations considered.

The averaged change in GPP is about -7.33 $GtC.yr^{-1}$ globally for the CO2 assimilation experiment. The corrections for the joint assimilations involving CO2 data is even lower: the mean global change are -1.07 $GtC.yr^{-1}$ for VI+CO2, -1.35 $GtC.yr^{-1}$ for F+CO2 and -1.98 $GtC.yr^{-1}$ for F+VI+CO2. For the F+VI+CO2 2-step experiment, the constraint on GPP is close to that obtained when CO2 data are assimilated alone (-7.70 $GtC.yr^{-1}$).

For the joint assimilations, the posterior errors on NEE is about 0.9 $GtC.yr^{-1}$ globally and about 0.3 $GtC.yr^{-1}$ for the three regions considered. The lowest posterior errors on GPP are obtained for the two experiments that combine the three data streams (about 0.09 $GtC.yr^{-1}$ at the global scale, and about 0.04 $GtC.yr^{-1}$ depending on the region). The values are close to the ones obtained with F+CO2.

Supplementary Text S7: Relative constraints brought by the different datasets with respect to PFTs and atmospheric stations

We performed the analysis of the influence of each data stream by discriminating the influence of each PFT for flux and satellite data, and each station for atmospheric CO2 concentrations (Figure S3 - experiment F+VI+CO2). For the flux data, the results are mainly proportional to the number of observations available (hence, the lower results are obtained for BorDBF, TeDBF and TrEBF, for which the number of assimilated data is about one order of magnitude lower than for the other PFTs; see § 2.2.1).

For satellite NDVI data however, the number of data is the same for each PFT. The discrepancies between PFTs is thus less pronounced than for flux data and related to the ability of the selected parameters to correct the phenology of each PFTs (constrained by the NDVI data). For TrDBF and C3GRA, the inability to correct the start of the growing season (Kpheno,crit, remains close to the prior values, as seen in Figure 3) may explain the lower contribution of these PFTs.

For atmospheric CO_2 data, the DFS is relatively well distributed across stations, with a mean value of 1.9 (range 0.19 – 14.5), in particular in the northern hemisphere. The higher values are found for a few southern hemisphere stations: Halley Station - HBA (6), Syowa - SYO (8.4), South Pole - SPO (11.9) and Cap Grim Observatory - CGO (14.5). Possible reasons for their larger impact may combine: a strong *a priori* model-data mismatch that is substantially corrected, ocean-driven concentration variations not well captured by the prescribed ocean flux but incidentally well corrected by remote land fluxes, etc.

Name	TropEBF^{F}	TropBRF^{VI}	TempENF^{F}	TempEBF^{F}	TempDBF^{F,VI}	BorENF^{F}	BorDBF^{F,VI}	BorDNF^{VI}	C3Grass^{F,VI}
Photosynthesis									
V_{cmax}	65 [35;95] *10*	65 [35;95] *10*	35 [19;51] *5.3*	45 [25;65] *6.7*	55 [30;80] *8.3*	35 [19;51] *5.3*	45 |25;65] *6.7*	35 [19;51] *5.3*	70 [38;102] *10.7*
$G_{s,slope}$	9 [6;12] *1*	9 [6;12] *1*	9 [6;12] *1*	9 [6;12] *1*	9 [6;12] *1*	9 [6;12] *1*	9 [6;12] *1*	9 [6;12] *1*	9 [6;12] *1*
T_{opt}	37 [29;45] *2.7*	37 [29;45] *2.7*	25 [17;33] *2.7*	32 [24;40] *2.7*	26 [18;34] *2.7*	25 [17;33] *2.7*	25 [17;33] *2.7*	25 [17;33] *2.7*	27.25 [19.2;35.2] *2.7*
SLA	0.0154 [0.007;0.03] *0.0038*	0.0260 [0.013;0.05] *0.0062*	0.0093 [0.004;0.02] *0.0027*	0.02 [0.01;0.04] *0.005*	0.026 [0.013;0.05] *0.0062*	0.0093 [0.004;0.02] *0.0027*	0.026 [0.013;0.05] *0.0062*	0.019 [0.009;0.04] *0.0052*	0.026 [0.013;0.05] *0.0062*
Soil water availability									
$H_{um,cste}$	0.8 [0.2;3] *0.47*	0.8 [0.2;3] *0.47*	1 [0.25;4] *0.62*	0.8 [0.2;3] *0.47*	0.8 [0.2;3] *0.47*	1 [0.25;4] *0.62*	1 [0.25;4] *0.62*	0.8 [0.2;3] *0.47*	4 [1;10] *1.5*
Phenology									
LAI_{MAX}	7 [4;10] *1*	7 [4;10] *1*	5 [3;8] *0.8*	5 [3;8] *0.8*	5 [3;8] *0.8*	4.5 [2.5;6.5] *0.7*	4.5 [2.5;6.5] *0.7*	3 [1.5;4.5] *0.5*	2.5 [1.5;3.5] *0.3*
$K_{pheno,crit}$		1 [0.7; 1.8] *0.18*			1 [0.7; 1.8] *0.18*		1 [0.7; 1.8] *0.18*	1 [0.7; 1.8] *0.18*	1 [0.7; 1.8] *0.18*
T_{senes}					12 [2;22] *3.3*		7 [-3;17] *3.3*	2 [-8;12] *3.3*	-1.375 [-11.4;9.4] *3.5*
$L_{age,crit}$	730 [490;970] *80*	180 [120;240] *20*	910 [610;1210] *100*	730 [490;970] *80*	180 [90;240] *25*	910 [610;1210] *100*	180 [90;240] *27.5*	180 [90;240] *27.5*	120 [30;180] *25*
$K_{LAI,happy}$	0.5 [0.35;0.7] *0.06*	0.5 [0.35;0.7] *0.06*	0.5 [0.35;0.7] *0.06*	0.5 [0.35;0.7] *0.06*	0.5 [0.35;0.7] *0.06*	0.5 [0.35;0.7] *0.06*	0.5 [0.35;0.7] *0.06*	0.5 [0.35;0.7] *0.06*	0.5 [0.35;0.7] *0.06*
Respiration									
Q10	1.9937 [1;3] *0.33*								
$HR_{H,c}$	-0.29 [-0.59;0.01] *0.1*								
MR_{c}	1 [0.5;2] *0.25*								
$K_{soilC,site}$	1 [0.5;2] *0.1*								
$K_{soilC,reg}$	1 [0.7;1.3] *0.1*								

Table S3: Prior value, interval of variation (in square brackets) and 1-sigma prior error (italic), of the optimized parameter. Except for those related to respiration, all parameters are PFT-dependent. The exponents F and VI associated to each PFT name indicate the availability of flux (F) and satellite (VI) data.

Figure S1: Monthly mean atmospheric CO_2 concentrations, for four stations (Mace Head - MHD (Ireland), Mauna Loa - MLO (Hawaii, USA) , Ryori - RYO (Japan), South Pole - SPO (Antarctic, USA)) over the period 2000–2009. The prior (blue) and the posterior (orange) model simulations are compared to the observations (black), and the corresponding RMSD is provided. The observation error is in grey. The dash lines correspond to the trend derived from the CCGCRV algorithm.

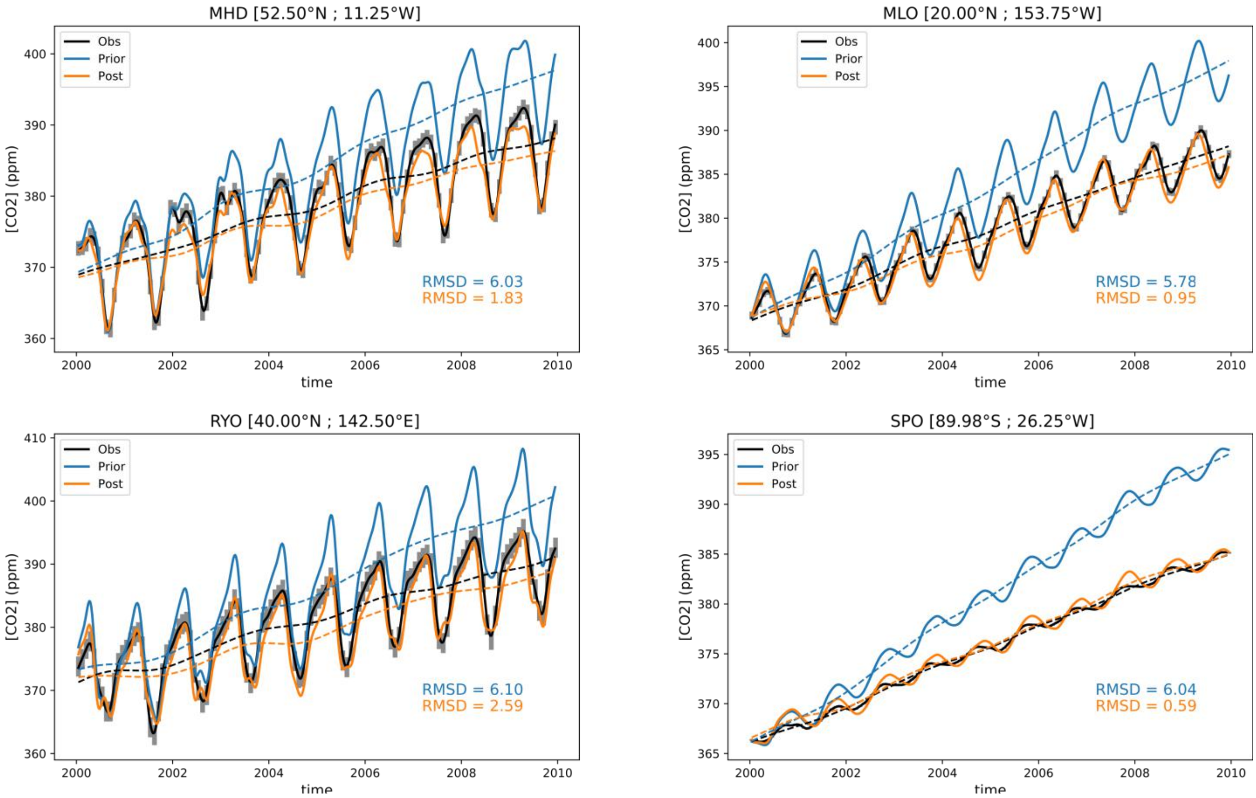
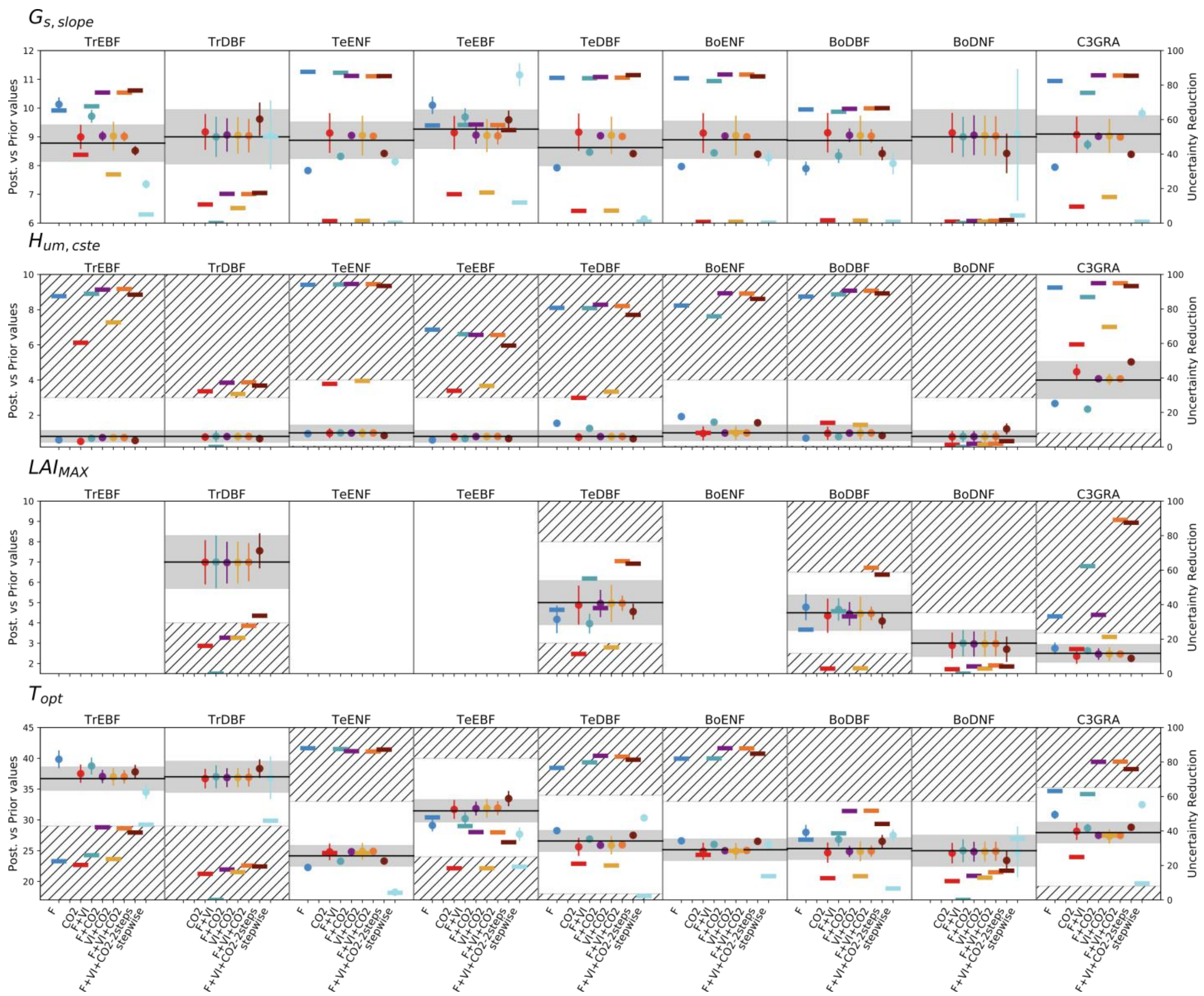


Figure S2: Prior and posterior parameter values and uncertainties for a set of optimized parameters (eight PFT-dependent parameters). The prior value is shown as the horizontal black line and the prior uncertainty (standard deviation) as the gray area encompassing it along the x-axis. For the PFT-dependent parameters, each box corresponds to a given PFT; empty boxes indicate that this parameter was not constrained for the corresponding PFTs. The white zone (non-dashed area) corresponds to the allowed range of variation. The optimized values are provided for each assimilation experiment ; the corresponding posterior errors are displayed as the vertical bars. Note that the prior values presented here are those used in this study, and not those of the stepwise (which are higher/lower for the photosynthesis and respiration / phenological parameters). For each assimilation experiment is also provided the uncertainty reduction (right y-axis) as the thick opaque horizontal bars.



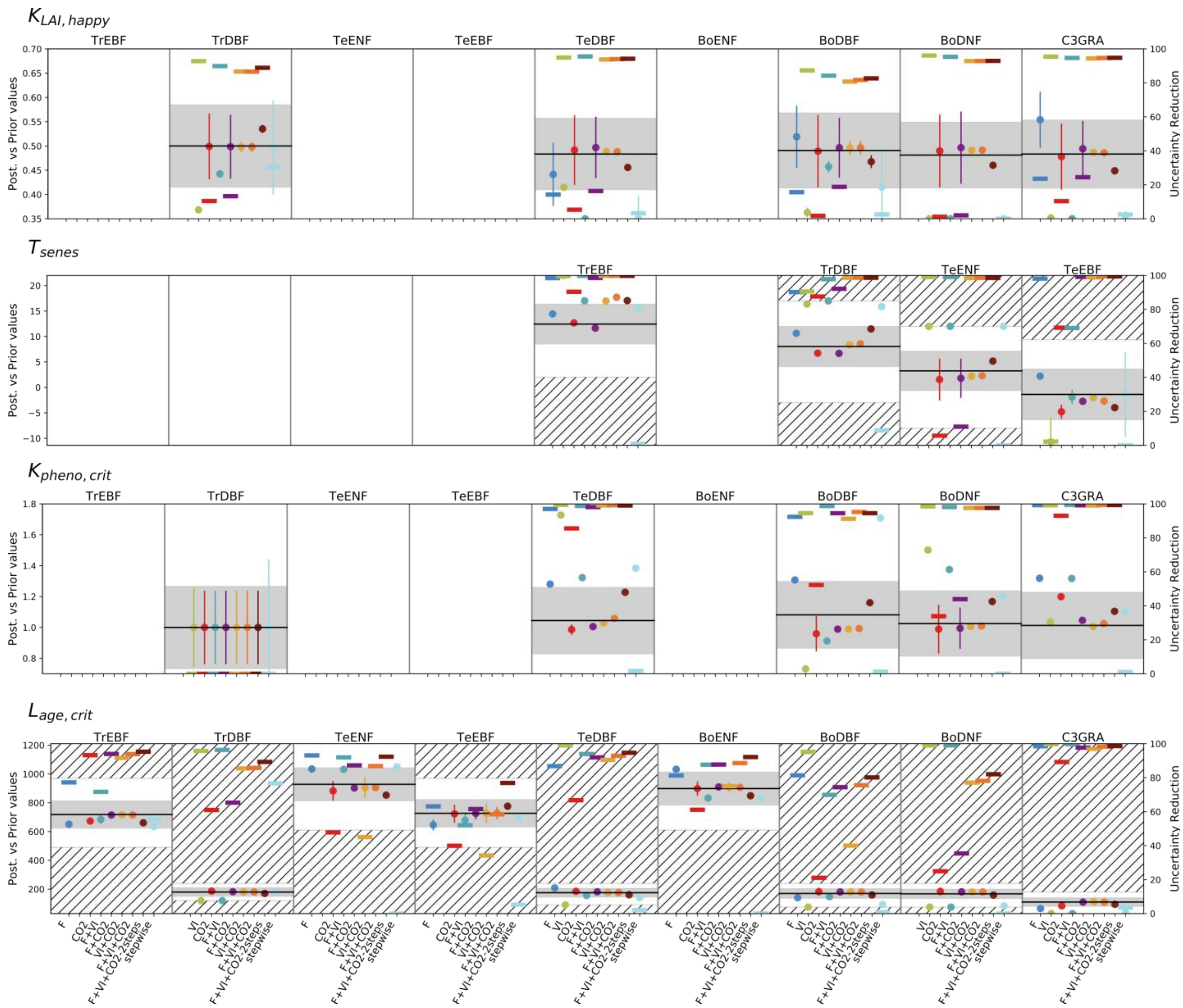
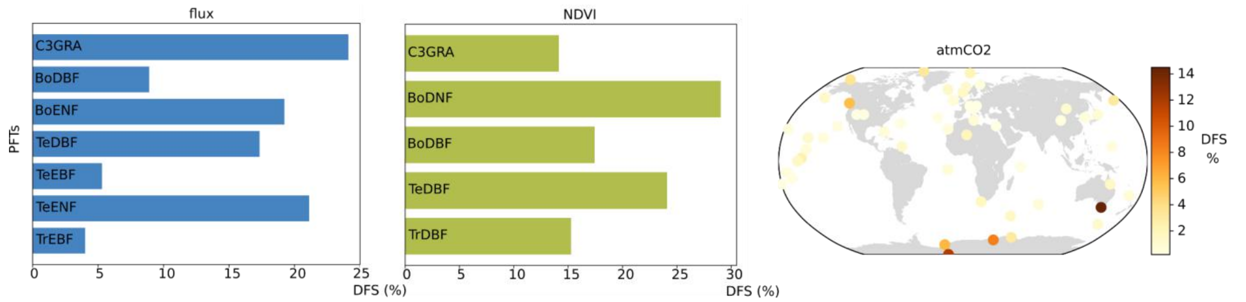


Figure S3: Relative DFS for the F+VI+CO2 assimilation experiment. For Flux and Satellite data: relative DFS per PFT; for atmospheric CO_2 data: relative relative DFS (contribution) of the different stations to the fit.



References

Carvalhais, N., Reichstein, M., Ciais, P., Collatz, G. J., Mahecha, M. D., Montagnani, L., et al.:. Identification of vegetation and soil carbon pools out of equilibrium in a process model via eddy covariance and biometric constraints. *Global Change Biology*, *16*(10), 2813–2829, 2010.

Kuppel, S., Chevallier, F., & Peylin, P.: Quantifying the model structural error in carbon cycle data assimilation systems*, Geosci. Model Dev., 6, 45–55, doi: 10.5194*. gmd-6-45-2013, 2013.

Kuppel, Sylvain, Peylin, P., Chevallier, F., Bacour, C., Maignan, F., & Richardson, A. D.: Constraining a global ecosystem model with multi-site eddy-covariance data. *Biogeosciences*, *9*(10), 3757–3776, 2012.

MacBean, N., Maignan, F., Peylin, P., Bacour, C., Bréon, F.-M., & Ciais, P.: Using satellite data to improve the leaf phenology of a global terrestrial biosphere model. *Biogeosciences*, *12*(23), 7185–7208, 2015.

Peylin, P., Bacour, C., MacBean, N., Leonard, S., Rayner, P., Kuppel, S., et al.: A new stepwise carbon cycle data assimilation system using multiple data streams to constrain the simulated land surface carbon cycle. *Geoscientific Model Development (Online)*, *9*(9), 2016.

Thoning, K. W., Tans, P. P., & Komhyr, W. D.: Atmospheric carbon dioxide at Mauna Loa Observatory: 2. Analysis of the NOAA GMCC data, 1974–1985. *Journal of Geophysical Research: Atmospheres*, *94*(D6), 8549–8565, 1989.