Supplement of Geosci. Model Dev., 17, 997–1021, 2024
https://doi.org/10.5194/gmd-17-997-2024-supplement

Geoscientific Model Development
Open Access
EGU

Supplement of

AgriCarbon-EO v1.0.1: large-scale and high-resolution simulation of carbon fluxes by assimilation of Sentinel-2 and Landsat-8 reflectances using a Bayesian approach

Taeken Wijmer et al.

Correspondence to: Taeken Wijmer (taeken.wijmer@univ-tlse3.fr) and Ahmad Al Bitar (ahmad.albitar@gmx.com)

S1 details on the BASALT implementation

S1.1 Computation time scaling regarding LUT size, number of pixels and zones

In AgriCarbon-EO, MCMC-based approaches are computationally inadequate because they would impose dependent iterative procedures over a very large number of inversion problems. This is a common feature of iterative methods when applied to a large number of inversions. For example, considering an iterative approach, the total computation time f(e) can be approximated.

$$f(e) = e \cdot N \cdot (t_{sample} + t_{run} + t_{eval}) \tag{S1}$$

Where e is the number of entities, N is the number of iterations, t_{sample} is the time to make the sampling, t_{run} is the process-based model run time, and t_{eval} is the time for each objective function evaluation. Given that the process-based models in AgriCarbon-EO are parsimonious, which entails relatively low complexity inversion problems, a non-iterative Bayesian NIS method is applied. NIS approximates the posterior distribution with a vector of weighted parameters or variables. This method has been applied to crop model inversions (**?**) but not to the assimilation of satellite imaging. Actually, when the processed entities are numerous for a similar set of forcing inputs (*i.e.* weather grid), the iterative inversion produces a considerable amount of common solutions in the explored parameter space (*i.e.* samples). Those common solutions hint at the possibility to use a LUT for each group of entities that share a common forcing and evaluate the relative likelihood of each LUT entry. Thus, if we consider that the number of samples in the LUT is n, the computation time g(e) of the new solution changes.

$$g(e) = n \cdot (t_{sample} + t_{run} + e \cdot t_{eval}) \tag{S2}$$

When the number of entities to inverse is large ($e.N >> n$) and ($t_{eval} << t_{run}$), this approach requires fewer simulations and lower time per entity. In the case of AgriCarbon-EO, the runs are done over tens of millions of entities, the evaluation time based on log-likelihood is extremely low, and since the scheme entails statistically independent samples runs are vectorized which reduces t_{run}. The solution presents one drawback in terms of higher memory overhead to store the LUTs compared to an iterative approach

S1.2 Likelyhood implementation

To gain computational speed, in AgriCarbon-EO, the loglikelihood is reformulated in a tensor form Eq.(S3) that uses the inner products of depth-2 arrays *i.e.* simple and efficient matrix product. This allows us to perform likelihood computations at a decent speed with the numpy library.

$$logL_{i,j} = -\frac{1}{2}(\sum_{o}(log(2\pi\sigma_{j,o}) + \frac{\mu_{j,o}^2}{\sigma_{j,o}^2} \otimes 1_i) + \sum_{o}(v_{i,o}^2.\frac{1}{{\sigma_{j,o}}^2}) + \sum_{o}(v_{i,o}.(\frac{\mu_{j,o}}{\sigma_{j,o}^2})) \tag{S3}$$

S1.3 Figures

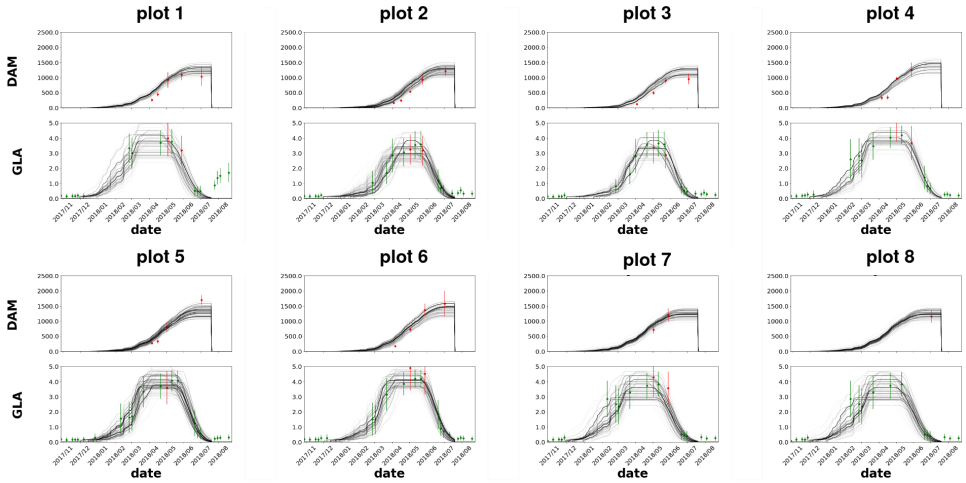

Figure S1. DAM and $GLAI$ time series at the location of the ESU measurements used in Figure 4. The observations of DAM and $GLAI$ are represented in red, the satellite $GLAI$ values in green, and the ensemble of model simulations in black with transparency relative to the likelihood of each element of the ensemble.

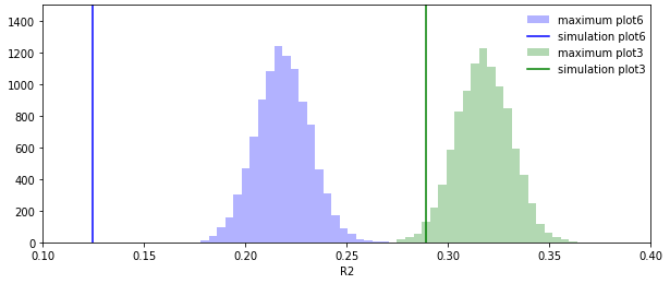

Figure S2. This figure illustrates the maximum R^2 expected for simulations if the combined yield map measurement error and the simulation error have a standard deviation of $1\,\mathrm{t\,ha^{-1}}$ and the R^2 obtained With AgriCarbon-EO. The distributions of maximum R^2 computed for 10000 repetitions of the comparison between the yield maps and the same maps with an added Gaussian noise of $1\,\mathrm{t\,ha^{-1}}$. The R^2 of the simulations are represented with vertical lines.

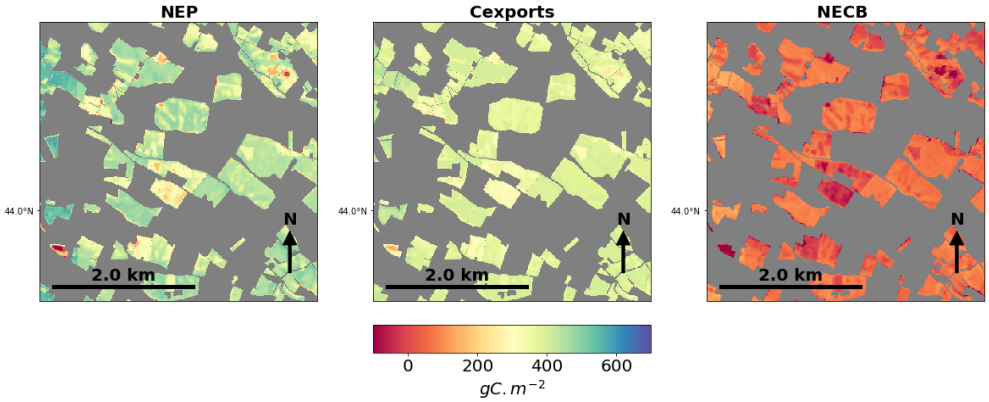

Figure S3. From left to right, NEP, Cexport (yield), and NECB for the winter wheat fields for the 2016-2017 cropping season. NEP and NECB have been represented with the soil-oriented convention to be able to compare carbon fixation to exports. In B. the histogram of the posterior mean and standard deviation of the same variables respectively on top and bottom. This figure is a zoom on the maps in Figure 6 that illustrates the retrieved inter and intra-field variability of our variables of interest around the fields where yield maps were acquired.