

Interactive comment on “Reviews and syntheses: Systematic Earth observations for use in terrestrial carbon cycle data assimilation systems” by Marko Scholze et al.

T. Kaminski (Referee)

thomas.kaminski@inversion-lab.com

Received and published: 12 April 2017

This manuscript presents several observational data streams that provide useful constraints in a Carbon Cycle Data Assimilation System (CCDAS). Such a contribution to the special issue covering the observational aspects, is very useful as it provides information that is complementary to the remainder of the articles in the special issue, which present application examples, overviews, or methodology. It is also reasonable to focus on a few Earth Observation (EO) data streams that were not covered in previous review articles on the subject. And it is reasonable not to focus only on direct observations of the carbon cycle but also address data streams such as soil moisture, which, through the process model, acts as an indirect constraint on the carbon cycle.

[Printer-friendly version](#)

[Discussion paper](#)



This special issue is one of the outputs of a working group of the international space science institute's working group "Carbon Cycle Data Assimilation: How to Consistently Assimilate Multiple Data Streams", the other one was a summer school, and one intention of this special issue is to provide further reading material to the students. In this context, the manuscript offers the right level of detail as a starting point, with many references to further material.

This is a good paper. My following comments are all minor and attempt

- to render the manuscript more useful for the target readers,
- to focus on the overall subject of the ISSI group and the special issue "consistent assimilation of multiple data streams", and
- to provide links to the other contributions.

The authors should address those they find useful.

1. Model-data fusion and data assimilation: L 37 states that both terms mean the same. Is this true? If yes, I suggest to keep one of the two for the rest of the manuscript instead of switching between the two. If not – maybe because by model-data fusion we could also understand some blending of observations with pre-computed model output – then be more precise in the definitions here and below use the appropriate term depending on context.
2. L 47: "new observation": maybe rather "new data stream" or "new type of observation"
3. L64: "In contrast to Ciais et al. (2014), who focus on carbon-cycle observations, we focus here on any kind of relevant observational data to be (potentially) assimilated in a terrestrial carbon cycle data assimilation system (CCDAS). In

Printer-friendly version

Discussion paper



a CCDAS the observations are used to constrain the underlying model (i.e. to move model output quantities closer to the observations and reduce their posterior uncertainties) usually by parameter optimisation.” This formulation could be improved. “any kind of” could be dropped in the first sentence, and the second sentence could read (for example): In a CCDAS non-carbon observations can be exploited to constrain the simulated carbon cycle indirectly through the relations implemented in the process model. Such observational constraints act by ruling out combinations of the unknowns in a CCDAS (typically a combination of process parameters, initial- or boundary conditions) that are inconsistent with the observations and thereby reduce uncertainties in the simulated output.”

4. L73: “Our focus lies on the terrestrial carbon cycle, because of the higher spatial and temporal variability in the net exchange fluxes and their associated higher uncertainties than form the ocean and anthropogenic components.” Maybe not true on all relevant scales. Maybe just drop the sentence, no need to justify the terrestrial focus in this context.
5. L101: in fact the weighting is in inverse proportion of the uncertainty, also appears below where Eq 1 is described
6. L103: either is maybe not appropriate?
7. L118: Maybe you want to put: “Here, we follow the notation as introduced by Rayner et al. (2016)” at the beginning of the subsection, i.e. before you start using their notation.
8. As we are dealing with assimilation of “multiple data streams” you could mention that usually each data steam requires its own observation operators, and in fact already here refer to Kaminski and Mathieu (2016/7), maybe even their octopus Figure. And for Eq 1 you could say that, for convenience of notation, now you combine all of them into a single $H()$.

[Printer-friendly version](#)[Discussion paper](#)

9. L122: Maybe drop this sentence. In fact the state are mixing ratios.
10. L133: “thus evolves”?
11. L136: “optimality” maybe you can find a better word? Maybe “adequacy”?
12. Figure 1 is confusing in several respects (prior info enters cost, inner/outer loop to be confused with NWP terminology, $U(o)$ not necessarily only a model output, cost function at minimum does not imply availability of A, etc ...). Maybe you just want to drop it with the two sentences that describe it?
13. L113-150 starting with “From Equation ...” could also be clearer, shortened or dropped (It does not follow from Eq. 1 that uncertainties are to be taken into account, but Eq. 1 follows from combining PDF descriptions of prior, observations, and model with a few simplifications, Mean and variance are not sufficient to characterise a multi-variate Gaussian, ...) Same holds for next paragraph (“assimilation problem is Gaussian” does not make sense; division by the matrix B is not straightforward...) Maybe just explain variables in Eq. 1 and then directly move to the paragraph starting with Rayner et al. (2016).
14. L161: The “and” between citations is missing (“citep” would have worked for multiple citations), same problem occurs a few times further down below.
15. L205: what about non-linear observation operators?
16. Section 3.1: It is good to introduce the different forms of errors. It would also be instructive to provide definitions of precision and accuracy.
17. L242: Is is worth mentioning that the scale at which we trust the model may be larger than a grid cell? L247: “In the case of satellite-based observations the representation error also includes errors in inferring a biophysical quantity from the photons measured at the sensor. We come back to this issue later.” I would think that such errors in the retrieval rather go into into the above two categories?

[Printer-friendly version](#)[Discussion paper](#)

18. L 255: “they affect the prediction of the optimal solution in the same way as” could maybe be replaced by “they have considerable impact on the solution. This is because of their influence on the weight of the respective observations in the cost function.” It is very good you stress this point. In fact you should take it up in the presentation of each data stream. So far it is only addressed in the XCO2 and the biomass sections.
19. L 263: What is inhomogeneous variance?
20. Section 3.2: Maybe add reference/web page of ICOS? Is is worth mentioning similar programmes outside Europe? “The measurements are designed”: maybe better “the network” or “the observing system”? L282: Paragraph may fit better into the beginning of section 3.3. Where you discuss the relevant observations provided by the sentinels, you are using our current perspicive, i.e. S1-5. You could make this clear, because in a few years time are reader could wonder why you don’t mention observations by S6 ... etc...
21. L 305 and 310: On L305 you write EO, then Earth Observation, then EO... something to be checked throughout ...
22. L344: For example Luke (2011) assimilates LAI.
23. Section 3.3.1: Is it worth to briefly explain how a total column value can be sensitive or insensitive to surface fluxes? L400: “)” should be “(“ L413: You could mention how the aggregation of errors to the 5 degree grid was performed. L435: Maybe update reference to latest version of the CCI CAR. L439: Is is worth mentioning planned XCO2 missions?
24. Section 3.3.2: L445: remove one “)”. L458: “closely follows the state of the vegetation” could be “is determined by the state of the canopy-soil system”L475: Disney et al. (2016) also compare two products. L 480: To simplify the sentence,

[Printer-friendly version](#)[Discussion paper](#)

maybe move the part in brackets up to the definition of L460. L503: “see 2” should be “see table 2”. Regarding correlation of uncertainty you might add on L506: after “periods.”: To reduce disk space, by default, JRC-TIP products are delivered without correlations among the uncertainties between individual variables, even though these correlations are available. An estimate of uncertainty correlation in space and time is not provided. The JRC-TIP products derived from MODIS (collection 5) broadband albedos minimise temporal uncertainty correlation as each collection 5 albedo value is derived as integral exclusively of observations over non-overlapping 16-day periods.

25. Section 3.3.3: L 514: “directly related” In the context of data assimilation, is it worth mentioning that there are complex processes which require complex models as observation operators for SIF, in order to extract the maximal benefit from this data stream? L 524: “lies” could be “relies”? L 530: Why is the simplicity of the forward model related to the fact that least squares is applied, which might also work with complex models? L 540: Does “a compromise” make sense here? Isn’t it rather the definition of the grid size that is determined by a compromise and the number of retrievals then just a function of this choice of grid size (plus the other factors mentioned)? When discussing the spatial and temporal sampling, it might be instructive to mention the variability in time as you do it in space.
26. Section 3.3.4: L600: “cost of the”: maybe better “cost of long” L652: “;5.” L658: “to improve the model’s hydrology” in fact in a CCDAS we are after the indirect constraint on carbon, so this restriction may not be needed here?
27. Conclusions: L724: “observational characteristics of the observational data” maybe you meant error or uncertainty characteristics? L730: “correlations”: “uncertainty correlations” or “error correlations” L732: “For example, while FA-PAR data constrain mainly the phenology component of a terrestrial carbon cycle model, soil moisture data, in contrast, constrain the hydrological component,” see

[Printer-friendly version](#)[Discussion paper](#)

above regarding the indirect constraints. You probably do not need to write this. In fact FAPAR can provide an important constraint on hydrology (see Kaminski et al., 2012)

28. Fig 3: I'd suggest to go for a 6 panel figure, the four maps are tiny; Better use degree symbol in caption.
29. Table 1: I suggest to replace "parameter" by "variable"
30. Table 4: You could add wave lengths to the bands, for many colleagues the band names don't mean anything.

There are quite a number of typos. Many of them (e.g. "Reflectamce-based" or "as-sessemt" or "observeing") can be detected by a spell checker...

[Printer-friendly version](#)

[Discussion paper](#)

