Biogeosciences Discuss.,
doi:10.5194/bg-2016-237-RC2, 2016

Biogeosciences
Discussions
Open Access

Interactive comment on "Reviews and Syntheses: Flying the Satellite into Your Model" *by* Thomas Kaminski and Pierre-Philippe Mathieu

Anonymous Referee #2

Received and published: 6 September 2016

* General comments:

- This manuscript is largely technically correct and appropriate for Biogeosciences. The following comments are submitted for the consideration of the Authors; they aim at improving the readability and impact of this paper.

- The Authors may want to clarify the intended audience, and then to fine-tune their manuscript to provide added value for that particular audience. Indeed, it is likely that those readers who are fully familiar with model inversion and data assimilation will clearly understand the current version of the paper, but may not learn much from such a generic presentation. On the other hand, someone who has no background whatsoever in the subject matter may find it difficult to benefit from the manuscript, due to the idiosyncrasies noted below.

Printer-friendly version

Discussion paper

- It may also be helpful to revise the paper with a view to homogenize its various parts, and better link the technical information that is provided to an overall (partly missing) context. For instance, Sections 1 and 2 discuss 'observation operators', state variables and other concepts, but do not say much or anything at all about assimilation and retrieval (other than the rather enigmatic statement on Line 26). Section 3 then jumps into these latter methods, but does not explain why they are needed in the first place. Since models have been adjusted to data sets for centuries, way before "assimilation" was in vogue, there is a logical or thematic gap here that only specialists knowledgeable with the field will be able to leap over...

- In the same vein, the manuscript would gain from a more consistent use of mathematical symbols. It is counterproductive, in such a general paper, presumably aimed at a large audience, to keep switching notations or to assign different meanings to a particular symbol along the way. The paper also makes extensive use of acronyms, in particular to designate space instruments, but only a few are explicitly expanded. A simple approach to address this issue may be to provide the URL to an appropriate web site where further information can be found.

* Specific remarks:

- The title of the paper is somewhat dubious: It is not the satellite that flies into the model, but a satellite model (actually an observation operator) that is merged into another model. The drive to get a "catchy" title is understandable, but this one may not be particularly successful, or representative of what is actually discussed. Of course, this is a purely personal impression and a minor issue...

- Line 27: Most models require not only the specification of initial conditions but also boundary conditions, where empirical evidence plays a crucial role too.

- Lines 29-43: This text fragment is very clear about the difficulty of exploiting empirical data into a model that works on different space and time scales and resolutions. Yet, none of the subsequent discussion appear to refer back explicitly to these important statements, for instance to explain how in practice the observation operator actually bridges the gaps between the instrument specifications, the observational protocol, and the modeling constraints. Again, depending on the expected audience, it may be pertinent to provide some more concrete information about the practicality of implementing such an approach.

- Lines 55-85: Section 2.1 is supposed to provide definitions of key concepts, but turns out to be rather obscure. For instance, the concepts of "state", "state variable", "state vector", "state space" are used without any explanation or context. Similarly, Section 3 discusses "model state", "observed state", etc. It may be useful to specify whether these expressions apply to the actual system under observation, or to the computer-simulated model. In any case, if the jargon of thermodynamics and systems dynamics is to be taken for granted, much of the rest of the paper might be of limited interest to those specialists. On the other hand, a clear introduction to these essential ideas could prove beneficial to readers unfamiliar with these concepts.

- Lines 65-67: The previous point is further highlighted by the unfortunate confusion between the state variables and prognostic variables. The former uniquely define the current "state" of the system, whether it evolves or not, and whether there is an attempt to predict its evolution or not. Prognostic variables are those that are forecast by a time-dependent model. These sets may be, but do not need to be identical. Clarity of mind is all the more important in this case because the general context of the discussion relates to the time evolution of complex systems such as the climate, or the carbon cycle, while the measurements obtained from satellites and assimilated through observation operators largely interpreted as repeated, but instantaneous snapshots (i.e., static processes), given the speed of propagation of electromagnetic waves compared to the rate of evolution of the system of interest.

- Lines 69-70: Similarly, the phrase "we must arrange for a change of the state" could be potentially ambiguous, as the actual state of the system is (usually) not changeable: instead, the "state variables" that describe the state of the simulation model are modified to reduce the "distance" between the simulated state (of the model) and the measurements, which in principle describe the state of the actual (observed) system.

- Figure 2: The right side of the graph is truncated: either move or re-design the whole graphics.

- Subsections 2.2, 3.2 and 3.3 enumerate multiple examples or applications that make use of observation operators, data assimilation techniques and retrieval methods. Such lists convey the message that these techniques are indeed exploited in a variety of fields, but do not really contribute to a better understanding of the design, development or functioning of such a tool. The point is not to delete these sections, but to clarify their purposes and added value: if they are meant to be a review of the field, then they may need to be beefed up. But if the intent is only to indicate that these techniques are widely used in a range of disciplines, then shorter sections or pointers to the literature may suffice.

- Line 121: Please note that both older (2004) and more recents (2013, 2015) publications have appeared on RAMI: see http://rami-benchmark.jrc.ec.europa.eu/HTML/RAMI-IV/RAMI-IV.php

- Line 154: If the purpose of the paper is to popularize advanced concepts as hypothesized above, then it might be appropriate to remark that the first term of Equation 2 is basically an expression of the least squares fit, which would be familiar to a broad range of readers. Similarly, the actual role and purpose of the second term of Equations (2) and (3) should be explained in more detail: Why is the first term insufficient? Could one rely on the second term only? Prospective users would likely gain from an understanding of these overall strategic questions before adopting these "advanced" methods.

- Lines 155-156: This enigmatic discussion about using x instead of z may be either pedantic or confusing: the symbols used in equations are irrelevant, as long as they are used systematically and coherently. Changing conventions in the middle of the paper is unjustified, especially given the minimalist use of the symbol z anyway. Please use the simplest set of mathematical symbols throughout the paper.

- Line 369: Delete the extra "around" near the end of the line.

- Line 400: What exactly is the implication of the phrase "because the dimension of the control space is large"? Is there a choice to minimize with respect to any other variable? Or would a different variable be chosen if the dimension of the control space were smaller, whatever that means?

- Line 405: It is not immediately apparent why Equation (9) is the gradient of Equation (2). It may be useful to be somewhat more explicit, or to point the reader to a more detailed (preferably publicly available) source.

- Lines 421-423: This statement about the limited accuracy of "the above listed algorithms" may need a bit more substantiation: indeed, the more advanced methods described here also have a limited accuracy, so the issue revolves around demonstrating that the uncertainty associated with AD is always lower than that of other approaches. While it is true that "incorrect gradient information will slow down or prematurely stop the iterative minimisation of J", why would those other methods systematically yield incorrect, or "less correct" gradients?

- Lines 423-424: Similarly, the statement concerning "the computational cost of this approximation grows linearly with the length of the control vector" may be true, but needs to be evaluated against a similar statement about what controls the cost of the AD method. Again, these claims may be correct, but they should be substantiated.

- Line 499: The phrase "maximum of the simulated area and 0 produces are step in the derivative at 0" is confusing or ill-stated.

- Line 502: Similarly, the phrase "describe to replacement" is odd.

- Line 520: The claim that "EO products can only be accessed by Earth system models via suitable observation operators" may be exaggerated: clearly, many users of EO products carry on without relying on, or even knowing about, observation operators. What may be more appropriate is to discuss why and to what extent such operators and the associated methods of assimilation/inversion provide more satisfactory results than traditional or earlier methods. An effective way to achieve this goal is to demonstrate the drawbacks that may arise when exploiting remote sensing data without relying on such advanced techniques.

I hope these comments may be helpful in updating the manuscript.

Interactive comment on Biogeosciences Discuss., doi:10.5194/bg-2016-237, 2016.