Author Response to Reviewer 2

Virtual joint field campaign: a framework of synthetic landscapes to assess multiscale measurement methods of water storage

Till Francke et al.
Geoscientific Model Development Discuss., doi:10.5194/gmd-2024-106

RC: ***Referee Comment***, AR: *Author Response*, ☐ Manuscript text

Dear Reviewer 2,

thank you very much for your positive response, and for the time and effort spent to examine the manuscript and the data set.

Your comments are very helpful in improving the quality of our manuscript. Please find a point-by-point reply below on how we intend to implement them. If you feel that some of these changes would not satisfy the needs you indicated, we would appreciate further advice on these matters.

Kind regards,
Till Francke (on behalf of the author team)

Comments and responses

RC: ***[...] The paper is well written and structured but there are a few issues that need addressing prior to publication – I don't believe any are particularly difficult to address.***

AR: We are grateful for the positive evaluation of the manuscript and the effort invested by both reviewers.

RC: ***The abstract would benefit from a sentence setting out the application context (e.g. soli moisture sensing). This is very nicely explained the introduction but the abstract starts on, for me, a rather technical note.***

AR: We will modify the abstract to better reflect this aspect as such (changes in *italics*):

> The major challenge of measurement methods *of environmental variables* beyond the point scale is their complex interpretation in the light of landscape heterogeneity. For example, methods like cosmic-ray neutron sensing, remote sensing, or hydrogravimetry are all able to provide an integral value on the water storage, representative for their individual measurement volume. [...] The present study demonstrates virtual observations *of water storage* with Cosmic Ray Neutron Sensing, Hydrogravimetry, and Remote Sensing in three exemplary landscapes.

RC: ***I appreciate the idea of the vJFC is to be quite generic (sentences around line 100), but I wonder if the title of the framework could be more specific to water storage applications? I think this would more accurately reflect what the code does at this stage – happy for you to argue against this if other applications are int eh pipeline.***

AR: Indeed, although the idea of the vJFC is quite generic, the current manuscript clearly focuses on applications for water storage. However, already the included examples also allow considering other target variables as well: both CRNS and remote sensing observations can also be used to infer biomass. We have indicated this in the Methods and also in the Conclusion. Likewise, if e.g. virtual temperature observations were included, the estimation of evaporation might also be envisioned. As we have not further elaborated these ideas so far, we would like to stick to just briefly mentioning this option, but still keep the name of the framework sufficiently generic.

RC: ***My main concern about the paper is that there is not overarching explanation or justification for the choice of case studies. Some are more detailed than others and the conclusions are essentially written without references to the case studies. I appreciate the test cases are simply illustrations of a huge number of potential applications, but a rational should be given for the choices, especially as some but not all of the measurement types are presented for each test case. How do the test cases come together to demonstrate the codebase effectively? And on that point do you do any tests (e.g. unit tests) to check the code is working as expected?***

AR: A similar point was also critizied by reviewer 1, so we would like to repeat the reply here:

Our focus was on the presentation of the framework, not on specific case studies. We used one realization (hexland_tracks) to demonstrate the applicability *across different sensor types*. As such, the other two case studies merely underline the versatility of the framework *across different landscape realizations*, specifically in representing terrain or resembling non-artificial landscapes.

We will clarify our focus on the framework and the illustrative and selective nature of our case studies and support this with the following table:

Table 1: Overview realizations

realization	**description**	**generated virtual observations**
hexland_tracks	synthetic landscape with max. contrasts	CRNS, remote sensing, hydrogravimetry
sierra_neutronica	synthetic landscape with high-relief	CRNS
agia	realistic Mediterranean landscape	CRNS

This table should help to illustrate how "test cases come together to demonstrate the codebase effectively", as phrased by Reviewer 2. Concerning tests of the code, the package includes a dedicated visualisation functions (which also served to create the illustrations in the paper). We will add a sentence to section "Scripts and external model code" to indicate this option for verification.

RC: ***I'm also not convinced I properly understood section 2.2.2 (L142). When merging if a compartment properties are replaced how is that different to stacking. Could some form of visualization be added to help with understanding how the landscapes are built in 2.2.2?***

We will add a figure similar to the one below to illustrate the different modes of combining the compartments:

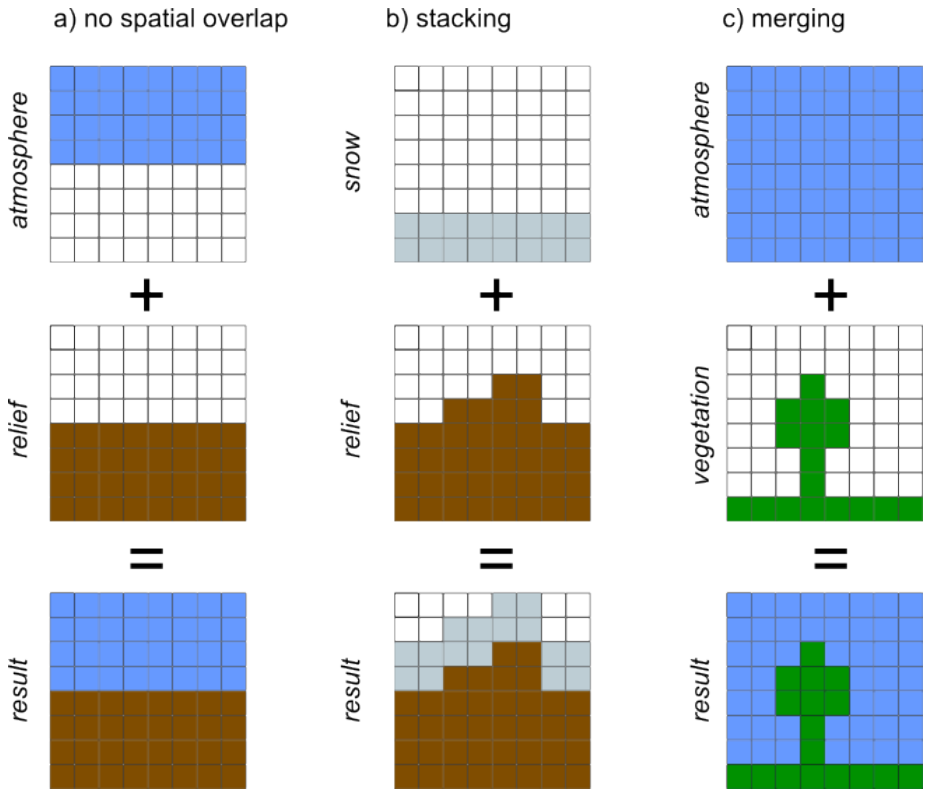

Figure 1: Simplified examples emonstrating the different modes of combining compartments into realizations

RC: ***I think GMD requires version numbers in the title?***

AR: The guidelines state " the title page must include the title (concise but informative, including model name and version number *if a model description paper*)". As we see our study clearly in the category "Methods for assessments of models", we have not included a version number in the title.

RC: ***L46: Check '(reference Welcome)'***

AR: Will be fixed.

RC: ***L373: This case study is introduced due to its importance for all three observation types, but only CRNS is presented. Why not present results form all three instrument types? At the very least a clear justification for not doing this is needed, especially as the justification is currently broader than what's presented. If the results are not particularly interesting for some reason could they be included nevertheless in a supplement? I appreciate it's likely not necessary to go into great detail about results form every test case, they are primarily illustrative, but a rational is needed for the choices about what to present (see my main comment). To be clear, I'm not advocating for more test cases and sensor examples, but the explanation of why various virtual sensor examples have been chosen needs to be much stronger.***

AR: We appreciate the reviewer understanding the illustrative nature of the examples. In addition to our reply on

Reviewer 2's respective main comment (see above), we will strengthen the justification of the selection in the beginning of section 3.

RC: ***L485: When 'relevant the aspects that merit further analysis' are presented these should be linked with the case study where they emerged when relevant. This basically links in with my main critique of the paper – the purpose the test cases and rational of their choice is not well summarized and then not well used to support the conclusions. The conclusions could disaggregate between perceived applications and those illustrated by case studies as a way to link the case studies and conclusions.***

AR: Section 4 is entitled "Conclusions and outlook". Due to the illustrative nature of the case studies and the focus of the paper on the framework rather than single aspects, we did not intend to draw substantial conclusions from (very briefly analyzed) results of the case studies. Instead, the mentioned list rather focuses on the "Outlooks", i.e. the envisioned scientific questions that can be further elaborated with the presented examples. However, we will also annotate the existing list of bullet points with abbreviations to indicate which case studies could be used to address the raised questions.

RC: ***L510: I'd like to see the limitations around assuming no measurement error introduced before the case studies, and certainly not in the conclusions (apologies if I missed this earlier in the manuscript). I was thinking about this issue while reading the case studies.***

AR: It is correct, that the presented examples do not consider any measurement errors. However, this is purely a decision made for the sake of keeping the examples simple. It is, by no means, a limitation of the concept of the vJFC. On the contrary, the amount of error and its effect can be systematically analyzed with this framework. We will clarify this in the respective paragraph.

Concerning the position of this aspect in the text, we intended to start with the bullet point list as related to / derived from the case studies. After that, some more general ideas, irrespective of the case studies, are presented. We will improve the wording to make this clearer.

RC: ***L517: Computational expense is often mentioned as a barrier, but I don't think estimates of the computational expense are ever given. This would be very useful practical information for anyone using the package. Do I need a HPC system for this or is this expensive in a single computer context? As with the measurement error limitations this needs presenting before the conclusions.***

AR: We agree. While the requirements for the employed radiative transfer model are relatively modest, the neutron and hydrogravimetric simulations require computational power of several single-CPU-days per realization. We will add this information to suitable section.