Dear Editor, Dear Referee,

We would like to thank you for further considering our manuscript and for the comments and suggestions made on our revised manuscript.

We believe that most of the concerns and comments from the reviewer are due to shortcomings in our explanation of the context and objectives of the modelling framework we describe - which aims to provide a blueprint and infrastructure for the future of the land components featured in Earth system models, so that they are more modular and more suited to simulate the terrestrial water cycle.

This is why the main focus in the revised version of our manuscript has been to clarify these objectives, to detail the limitations in the current land components of Earth system models, and to explain and describe how the framework is addressing those limitations. In addition, we have replied and addressed, where appropriate specific reviewer comments as detailed below.

We hope that our revisions address any misunderstanding and meet the expectations of the Editor and the Referee.

Thibault Hallouin, on behalf of the co-authors.

Nomenclature:

RXCY – Referee number X Comment number Y
AR – Authors' Reply
PXLY – Page X Line Y

Referee #3

General issues

R3C1: The manuscript "UniFHy v0.1.1: A community modelling framework for the terrestrial water cycle in Python" presents a code framework for the coupling of distributed large scale models of the hydrological cycle – at least, this is how I understand the manuscript.

AR: No, this is a framework which allows decomposing land surface models developed for Earth System Models into only a few components. The proposed components correspond to ontologically well-defined entities. As they also correspond to disciplinary divisions, clear definitions of interfaces are essential so that microscale meteorologists understand the needs of hydrologists, or soil biochemists. This also explains why the subdivision of the land system needs to be as parsimonious as possible.

R3C2: The scope and the topic of the manuscript is not well defined and the basic motivation to create the code is not well explained. The introduction mentions a couple of existing modular modelling approaches but they do not really categorize the existing approaches and, most important, the introduction misses a section explaining what problems are not solved with the status quo and which benefits are created by the presented software.

AR: The paper in its introduction explains the limits of "existing modular modelling" in Earth system dynamics and in Earth system modelling communities. We argue that an intermediate approach is needed. UniFHy demonstrates that such a coarser subdivision of the land system, suitable for Earth system models, is achievable.

R3C3: From the whole manuscript I understand that UniFHy is a coupling interface definition, like OpenMI, OASIS-MCT / Open-PALM etc. that does not include solvers and / or model equations like the cited modular Hydrology tool kits (FUSE, SuperFLEX, CMF, SUMMA and RAVEN) or LandLab. However, the interface is extremely limited in scope and it is not well explained, why such an interface is needed. My own experience is that any Python compatible API to model internals is sufficient for effective code coupling.

AR: It is exactly the ambition of the proposed interfaces to be limited in scope as they need to be implementable in OpenMI and OASIS. The solvers are internal to the three components proposed. It has to be recognised that a LSM is typically applied globally so that a large number of special cases need to be dealt with, thus general solvers are often unfit for this purpose.

R3C4: Discussions about model boundaries (eg. is root distribution subsurface or vegetation model component) is the most important part of model coupling from my experience. This discussion is missing in the manuscript and hindered by the interface. For my coupling studies such an interface would have been a hindrance and not an asset – even if Landsystem models are involved (CLM, LPJ-GUESS) (Bendix et al., 2021).

AR: Again this is internal to each of the three components proposed. The proposed interface can be implemented, and already partially exists in the LSMs the authors have contributed to (JULES, ORCHIDEE).

R3C5: The manuscript needs a complete rewrite to address the question, what unifhy really provides, what the status quo is missing. Large parts of the manuscript can be omitted (future development, case study in its current form), or need a complete rewrite focusing on the main questions (introduction, description of the framework, usage). I would therefore recommend to reject the manuscript in its current form.

AR: We believe that the clarifications brought to the introduction better places the objective of this paper within the state of the art. It should also better explain the innovation UniFHy provides in this effort to decompose LSMs into better defined components.

R3C6: The major problem is: the authors have not really decided which of two possible papers they want to write:
a) if the m/s is about the interface only, the need and the practical features of the interface and existing helper functions (eg. grid transformation, parallel memory allocation) should be presented. The status quo needs to be presented in a structured form
b) the presentation of the case study is to short and limited for a story of its own: The results are not discussed and the framework in its current form can be used for more variants then just the three presented.

The m/s is a mix of both papers, but none of the versions is presented in adequate form.

AR: This manuscript is presenting a new modelling framework. To this end, we believe that the reader needs to get a description of it, but also a demonstration that it actually can be used for what it is designed to do. This is precisely our intention with the case studies, and as mentioned by the reviewer, they are not meant to be a story of their own, this is why we are trying to keep it relatively short.

The framework is designed to allow the decomposition of the terrestrial water cycle into components such that (a) alternative science components can be compared, and that (b) varied spatial and temporal resolutions can be used for each component. To demonstrate this, the case studies need to show the framework outputs using (a) different modelling components, and (b) different resolutions (hence 3 configurations). In addition, we believe that it is important to demonstrate that it can work in more than one location (hence three study catchments). This requires a brief description of the components and configurations used.

But, as mentioned in the manuscript *"This brief analysis of the results is used to demonstrate the potential of the framework to elicit the most suitable combination of components to simulate the hydrological behaviour of a given region"* (see P17L303-P18L305 in the reviewed manuscript). Showcasing outputs (hydrographs and spatialised streamflow simulations) appears to us as the bare minimum in view to demonstrate that it functions, however, if the reviewer would like for the comparative analysis with the performance metrics to be removed from the analysis, we are happy to consider it.

R3C7: There are also code quality problems, like the use of underscore as a prefix in the public (user / contributor facing) interface.

AR: It is correct that the use of leading underscores is a convention in the Python community to denote "private" attributes of objects. As the reviewer points out, there are two levels of interaction with the framework: as a user (to run simulations) and as a contributor (to develop modelling components). On the one hand, users are not made aware nor expected to interact with "private" attributes, as per the convention. On the other hand, contributors are expected to assign values to "private" attributes to declare their modelling component's inputs/parameters/constants/outputs of their own classes. They inherit from framework component classes, and as such, it is not going against the convention to assign values to their own component class' "private" attributes. These attributes are made "private" following Python's convention so that users do not interact with them, which would most certainly result in breaking the modelling component altogether. If the reviewer has other code quality problems to raise, we would be willing to consider them.

Specific issues

1 Introduction

R3C8: L40: CMF is not pure bucket style, but allows a wide range between physical and conceptual models – more physics than RAVEN (eg. Richards equation).

AR: Thank you for the correction. Given that the distinction in this paragraph was intended more about the granularity of the framework's building blocks, we agree that CMF belongs more with SUMMA and Raven as it is based on a finite volume method where process equations can indeed be replaced, as they form the model building blocks. The paragraph was rearranged accordingly (see P2L47-P3L52 in the revised manuscript).

R3C9: L43: "However, refactoring existing land surface models using these frameworks is not trivial" – citation needed. Counter examples for CMF coupling with CLM and LPJ-GUESS (Bendix et al., 2021). Using the proposed interface would be harder.

AR: This statement is made based on our experience, and experience collected from others (e.g. refactoring of CLM in SUMMA), so no particular reference can be included here. However, the sentence was rephrased to highlight that it is based on the authors' experience (see P3L52-43 in the revised manuscript).

R3C10: L44: OpenMI is something completely different from the modular hydrology toolboxes, it is more like the following paragraph

AR: As highlighted in the abstract, the framework is intended as a cross-community initiative, so we intentionally structured the existing literature per community of origin. Nevertheless, we do agree with the reviewer that OpenMI is closer to Earth System modelling frameworks in its philosophy, yet it was developed by hydro(geo)logists.

R3C11: L54: What is wrong with the status quo? Where are the gaps? Why don't you use an existing coupler or none at all?

AR: Current LSMs lack modularity and the entry-level to contribute to them is rather high and model-specific, which impedes comparisons of alternative processes from different LSMs. Frameworks used by hydrologists and Earth system dynamics are not fit for purpose because LSMs are components of ESMs and inherit the technical and computational requirements of ESMs which are not met by these frameworks. Existing modelling frameworks used for assembling ESMs feature such requirements, but they lack some specific needs of LSMs such as supermeshing or lateral fluxes crucial for hydrological processes. A new paragraph has been added to answer those questions (see P3L72-79 in the revised manuscript).

R3C12: L63: What is the "integrated coupling philosophy"? Which of the mentioned modular modelling approaches follow that also? What are the specific features of UnifHy?

AR: This term is explained earlier in the paragraph listing the existing efforts in the Earth system modelling communities (see P2L50-51 *"integrated coupling frameworks such as ESMF (Collins et al., 2005) or CPL7 (Craig et al., 2012), where existing modelling components require code refactoring to comply with a set of organising and interfacing requirements"* in the reviewed manuscript). ESMF and CPL7 are examples following this approach. UniFHy is an application of this existing approach to a new domain (i.e. the land system) with new interfacing requirements tailored to hydrology in the land system.

2 Description of the framework

R3C13: L85ff: Citation needed - why do you think a one size fits all solution ("skilfully yet pragmatically") exists? The "intentionally limited" "degrees of freedom offered" makes the proposed interface in my eyes anything else than "unified", but rather specialized for the purpose of this team and the models used in this study. What about energy or solute fluxes?

AR: The interfaces presented in the manuscript are the fruit of a large consultation with the land surface and hydrological modelling communities during workshops and follow-up discussions that will be summarised and made available separately (manuscript in preparation). As such, these interfaces are indeed specialised in subdividing the land system into finer components, based on the scope of such models and the processes typically taken into account in existing land surface models (e.g. ORCHIDEE, JULES, CLM). The models used in the case studies (i.e. Artemis, RFM, SMART) were subdivided as a subsequent step. The energy cycle is implicitly considered with the water cycle, even though thermodynamic fluxes are currently missing from the interface. The nutrient cycle is on the roadmap and is currently being defined separately.

R3C14: L98ff: Again, the term “integrated coupling approach” remains unexplained. The rationale for this structure is not explained

AR: See the author’s reply to **R3C11**.

R3C15: L136: “These contributions can be implemented purely in Python, but can also rely on Fortran, C, or C++ programs called by interface middleware.” From my experience, creating any interface is far from trivial. Using any interface for coupling is simple.

AR: We agree that interfacing compiled libraries with the Python framework is not trivial, but this does not make our statement any less valid. Indeed, the unit tests implemented in UniFHy do feature Fortran and C++ components wrapped in Python classes, which demonstrates that it is possible, and provides examples to follow in turn (i.e. using `numpy.f2py` and `cython`). The reader is made aware of this potential source for inspiration (see P15L256-257 in the revised manuscript).

R3C16: L154: “Contributors need not handle basic functionalities such as memory allocation nor input/output operations, as these are handled by the framework” Is this the main service of the provided software? How does the framework handle this? Why are the functions for memory allocation and input/output operations not explained? What are the limitations of the framework as is?

AR: As a general rule, given that the objective is to limit the burden of these tasks for the component contributors so that they can focus on the science aspects of the component, we are unsure of the relevance of documenting these aspects explicitly. That being said, regarding input/output operations, the contributor (and the manuscript reader) is made aware that the user is to provide NetCDF files whose data and metadata is following the CF conventions. Regarding memory operations, there is currently nothing special or innovative about our implementation worth mentioning, beyond the obvious care taken in avoiding unnecessary copies, memory leaks, and limiting the amount of data loaded in memory at once (through time slices). However, when the framework becomes ready to be used on distributed memory architectures (i.e. when domain parallelisation is implemented and compatible with the supermeshing approach), we do agree with the reviewer that the approach will be worth detailing.

3 Usage of the framework

R3C17: The whole section does not really follow a user story and is hard to follow. The code examples and descriptions seem to follow an example setup that is not yet described. The code examples lag explanations and comments.

AR: This section is structured as a step-by-step workflow that a user needs to follow, and as such we do believe that it follows a user story. Nonetheless, we have added a figure to provide an overview of the workflow to follow to set up and use the framework. We have also added paragraphs at the beginning of the section to better explain the content of the section (see P8L187-P9L194 in the revised manuscript), and more references between the text and the code examples. Together, we hope that these changes will contribute to make it easier to follow this section. As to the lack of code comments and explanations, please refer to the captions of the code snippets that we have expanded where appropriate.

4 Contribution to the framework

R3C18: This section would benefit of less code and more story.

AR: Providing less code appears undesirable as this would render the component invalid (i.e. it would not be a valid implementation accepted by the framework). So, we were only able to add a few extra elements to complete the story in this section.

5 Case studies using the framework

R3C19: The results are not discussed and the setup (and its rationale) is not well explained. Instead of focusing on the “unified” interface the m/s could present the results from this multi-model exercise as a scientific paper and mention, that more is possible using the interface.

AR: As already mentioned in the author’s reply to **R3C6**, we believe that this section is required in this manuscript, and that is not meant to be a story of its own, but rather a demonstration of what is claimed previously. However, we agree with the reviewer that the rationale for the setup was lacking, so we have remedied this (see P15L280-284 in the revised manuscript). The results are already discussed in the reviewed manuscript (P16L293-P17L302), through some examples of comparisons of the performances of alternative configurations rather than a complete analysis of all configurations in all study catchments, which would be beyond the scope of this manuscript, and which would go against the request from the reviewer not to try to produce a two-headed paper (as expressed in **R3C6**).

6 Future developments

R3C20: This section is far too long for things not yet existing. A short outlook is appropriate but two manuscript pages concerning not yet existing features is clearly too much – the future is always the most difficult time to predict, as we modellers know.

AR: As recommended by the reviewer, this section was removed, and the short outlook suggested is appended at the end of the Conclusions section (see P21L354-P22L366).

References

Bendix, J., Aguire, N., Beck, E., Bräuning, A., Brandl, R., Breuer, L., Böhning-Gaese, K., de Paula, M. D., Hickler, T., Homeier, J., Inclan, D., Leuschner, C., Neuschulz, E. L., Schleuning, M., Suarez, J. P., Trachte, K., Wilcke, W., Windhorst, D., and Farwig, N.: A research framework for projecting ecosystem change in highly diverse tropical mountain ecosystems, Oecologia, 195, 589–600, https://doi.org/10.1007/s00442-021-04852-8, 2021.