Earth System Dynamics Discussions Article

Earth system modelling with complex dynamic human societies: the copan:CORE World-Earth modeling framework

Final authors' response

Jonathan F. Donges*, Jobst Heitzig*, Wolfram Barfuss, Johannes A. Kassel, Tim Kittel, Jakob J. Kolb, Till Kolster, Finn Müller-Hansen, Ilona M. Otto, Marc Wiedermann, Kilian B. Zimmerer, and Wolfgang Lucht

(Reviewers' comments cited in italics)

Response to Carsten Lemmen (reviewer 2)

1 General comments

This manuscript by Donges and colleagues introduces the core technology and concept behind a new software tool called "copan", that should serve as "a framework for developing, composing and running World-Earth models". The authors motivate the development of such World Earth Models (WEM) that encompass dynamic descriptions of both the anthroposphere as well as the Earth System, they contrast WEM to integrated assessment and Earth System models, they describe the concepts of the developed software package pycopancore and they show simple example applications of the software.

The contribution is within the scope of the Special Issue "Social dynamics and planetary boundaries in Earth system..." in Earth System Dynamics, although the preferred outlet for this kind of technical model description could also be Geoscientific Model Development. The novelty of the approach is the complexity of a World model combined with a stylized version of an Earth model; the innovation is in the open framework and theoretical embedding of the World Earth Model approach.

The paper is overall well written, but suffers from resilience theory and technical jargon, which should be reduced to address a wider readership.

We thank you for this overall assessment and will aim at making the MS more accessible by reducing jargon, especially in the introduction, which will be shortened in response to Mr. Dermody's comments, and by giving additional definitions where necessary.

Figures are appropriate but they are of mixed graphical quality and accessibility and should be improved on. Tables are appropriate throughout; code examples examples are useful but in need of better quality. The supplementary material is well presented and useful.

In the revision, we will aim at improving the (old and new) figures' and code examples' appearance.

The theory-laden motivation somewhat contrasts with the very technical model description. Reviewer one already remarked on the need for better embedding of these two major perspectives the manuscript assumes. I agree with that assessment, but for brevity I will concentrate in my detailed review below on other aspects of the manuscript. A major missing part is a description of how the presented copan:CORE framework fits into and operates with much of the existing

coupling and model infrastructures in Earth and Social sciences; claims to interoperability, modularity and flexibility remain unsubstantiated.

We realize that this has to be improved, see our responses below.

I recommend that this paper is published after substantial revisions.

We thank the referee for his overall encouraging assessment.

Title, Abstract and related parts of Introduction

Title. There is an inconsistency in the spelling of "modelling" right in the title. Also, consider to spell out WEM as World Earth Model without hyphens; carefully consider lowercase/uppercase for "Model" in WEM. Nowhere in the paper the authors motivate the naming "copan:CORE"; please add a sentence on this naming and add to a table of abbreviations, if any of this is an acronym.

Thank you for pointing this out. We will check all our spelling again carefully. The hyphen in "World-Earth" has become somewhat of a standard spelling, but we will reconsider it. We'll explain the naming "copan:CORE"; "copan" is the name of PIK's flagship activity for studying coevolutionary pathways, all our models are named "copan:XYZ", and "CORE" refers to the modeling framework which will form the core of our working group's model portfolio.

p1 l1ff. That first sentence "Possible future trajectories of the Earth system in the Anthropocene are determined by the increasing entanglement of processes operating in the physical, chemical and biological systems of the planet, as well as in human societies, their cultures and economies" is very debatable. "Possible" is redundant, the choice of Anthropocene (capitalized) possibly politically motivated, the word "determined" raises concern of confusion with "deterministic" approaches and the conjunctions are not well placed. If I may rephrase this, the "Anthropocene (sic!) is characterized by close entanglement between the Earth system and its physical, chemical and biological processes and the World system with its economic, social, and cultural interactions." And certainly there is no need for eight (!) citations to entanglement in the Anthropocene; possibly, authors who argue for entanglement in the anthropocene (minuscule "a") should be cited instead.
p1 l3ff. Second sentence "Here, we introduce the copan:CORE open source software library that provides a framework for developing, composing and running World-Earth models..." This sentence should foremost and first emphasize that this publication introduces a new term and concept, namely that of a WEM, and second that it also provides a software library for modeling such WEM. Also the definition of WEM as "social-ecological co-evolution up to planetary scales" does not agree exactly with the later definitions given in the manuscript. Please elaborate in the abstract on your term WEM, on the theoretical embedding and reduce the room given to technicalities.

We thank the referee for these insightful remarks and will carefully revise and shorten the abstract accordingly.

Introduction.

p2 l25ff Please provide a reference your historical examples. In the discussion of the "Tragedy of the Commons" it would not hurt to point to related works that make Ostrum's work operational in model simulations.

We will add such references on historical examples and modelling studies operationaliting Ostrom's framework.

p2 l34f I believe the term "planetary social-ecological system" needs more explanation. SES are usually understood as local in much of the literature, and as multiple instances that behave very different. Thus, also the implementation of SES mostly in agent-based models (as you mention yourself later in the introduction). Elaborate and contrast your "planetary" approach to the local SES. You might also consider to reduce usage of the term SES altogether in favor of your new term WEM to avoid this confusion.
p3l 7ff Congratulations on the choice of the term "World Earth Model". This is to date the best term I have yet heard to describe the type of model you've developed. I suggest to elaborate on how you come to this term, and to set it off from other terms including, but not limited to, SES and CHANS (Coupled Human and Natural Systems).

We will revise the introduction to clearly define and explain these terms and their interrelations and differences, particularly to clarify that some of them refer to real-world systems while other refer to models of the latter.

Blueprinting World Earth Models

p3 l6ff Please use precise language, do not "outline guidelines" or "address leading research questions". Check entire manuscript for this type of bloated wording.

We will revise the text to ensure a more concise and crisp language.

p3 l7ff For the definition of an Anthropocene you already need to say how it differs from the Holocene and other paleoclimatic stages. So the first half of question type 1 is circular. As for the second part "how might it alter the evolution", it is unclear what "it" refers to. Certainly the "Anthropocene" is not an actor (so it cannot alter) but a diagnostic term for the World-perturbed Earth. Please clarify.

Well spotted. We will carefully revise and clarify this sentence and others relating to the notion of the Anthropocene.

p3 l8ff Avoid general valueing statements like "disastrous" or specify; check entire manuscript for further occurences of such type. Avoid jargon here and explain all domain-specific terms.

We very much agree and will revise the text accordingly to avoid unnecessary jargon.

p3 l27 Here you use "framework" in the management sense, later you use (software) "framework" for the technical description.

You are right, this was an unsensible choice of term here. We'll rephrase this sentence, avoiding the word "framework". We will reserve the word "framework" for its software meaning in the MS and will state so in the text.

Then you both consider Netlogo as well as copan:CORE frameworks, but both are very different things. I would prefer to term NetLogo a modeling platform.

We agree since NetLogo provides a graphical interface and other features typical of a modeling platform.

The term "framework" is a difficult one, please try to use it consistently in only one sense (and explain that sense by giving your definition of a framework) throughout the paper.

We'll add a short definition of "modeling framework" similar to the one of "software framework" that can, e.g., be found on Wikipedia.

p3 l27 The "high degree of modularity and flexibility and coupling capabilities" is not substantiated. While there is some software modularity and role modularity (see my later comment), there is no effort made towards coupling capabilities in a more general sense (there is a statement later on interoperability with LPJml, see my comment below). There is also no elaboration of what you mean by flexibility.

We agree that our discussion of these aspects needs to be improved. We'll comment on modularity and coupling below; by "flexibility" we mainly mean the possibility to use various combinations of modeling approaches and levels of aggregation (i.e., on the individual, cell, social system, or global level), so that one might combine an ABM of a labour market at the micro-level (i.e., individuals) with a system of ODEs modeling a carbon cycle on the cell level (photosynthesis) and global level (ocean-athmosphere diffusion) and a system of implicit and explicit equations representing a multisector economy with perfect factor markets on the social system (e.g. country) level.

p4 l14ff I don't see how the stylized biophysical description in the WEM can help answer this question. Would we not need a "whole" WEM where both the Earth System and the Socio-cultural system are described process-detailed (ref your Fig 1)?

The simple example WEM described in Sec. 5 is not meant to be a candidate for a meaningful WEM that could be used to answer real research questions. It is given only to illustrate the features of the modeling framework that this MS is about. If a user of copan:CORE deems it necessary to represent certain processes in more detail than others to be able to answer some specific research question, she can develop a model component that does just that or that acts as a wrapper around an existing external model software implementing these processes (see our comments on coupling in the response to Mr. Dermody and below). Although this is not too relevant here, we however personally believe that the specific question we gave as an example of a research question, namely "How does climate change feed back on complex social structures and their dynamics?", *can* be studied to some extent by a model that has only a stylized biosphere. E.g., changes in global mean temperature can lead to economic damages and increased average mortality, which in turn can lead to changes in demographic structure and economic processes and eventually to changes in social coherence. This is not saying that we already have all the necessary model components or even the theoretical or empirical means to formulate these model components, but that if one had these then a stylized biosphere component might well suffice to perform useful studies.

p4 l25ff You argue that environmental and societal processes should be described on a similar level complexity, yet in Figure 1 you argue for a stylized description of the biophysical world. Please explain better or resolve this conflict between text and figure.

This is a valid point, we will carefully revise the text in current Section 2 and Figure 1 to resolve this apparent inconsistency.

As for your list of five characteristics of WEM, I suggest to give each item a short title. You might want to consult our modeling framework paper (see references, we had to argue for biological models on par with physical oceanography models and called this "equitability"). Others could be "nonlinearity" and "aggregation".

We agree with this suggestion and will add summary titles to the WEM characteristics, referencing also to your recent modeling framework paper in the same special issue.

copan:CORE WEM framework

p6 l22ff Your modularity is achieved through object-oriented programming. This is not enough to justify modularity as an eminent feature of your software. This is mere good software practice. Object-oriented programming then does not per se allow interoperability and dynamics coupling to other models, as you claim.

We believe this is a misunderstanding caused by sloppy wording in the original MS. Of course we do not claim that object-oriented programming automatically leads to either modularity or interoperability. We will make sure in the revision that it becomes clearer that the high degree of modularity is the result of very specific design choices (which we found to be easier by following an object-oriented software design pattern rather than, e.g., a functional programming one), such as using multiple inheritance to allow different model components to use the same entities and attributes.

To this end, much more (like coupling frameworks, data exchange standards, computational bridging infrastructures) are needed, all of which are absent from the manuscript. Please elaborate on the specific coupling solution to LPJml and to IMAGE to substantiate your interoperability claim.

As already hinted at in our response to Mr. Dermody, interoperability with LPJmL, IMAGE etc. follows from the flexibility to basically use any Python code whatsoever in a model components' process implementation methods, including any calls to external software in order to exchange data or call stepper functions etc.

p8 l14ff Consider making this list of process-types identical to the one found in figure 2

Perhaps another misunderstanding. There is no list of process types in Fig. 2 but a list of modeling approaches. While there are some one-to-one relationships between the latter and the former, e.g. the modeling approach of using ODEs is supported by providing a process type "ODE" implementing a system of ODEs, other modeling approaches will require several formal process types, e.g. the ABM and adaptive network approaches will typically require a combination of processes of the formal process types "event", "step", and "explicit". We will try to include a similar clarification into the revised MSP.

p9 l16ff It should also be the role of the "master" model to ensure interoperability with other modeling frameworks, of which you make no mention.

We agree that both the "*base* model component" (implementing the most basic entity-types and relationships every copan:CORE model *must* have) and the "master *data* model" (a repository of entity-types and attributes model component developers *may* use) should aim at supporting as much interoperability with other models as possible. copan:CORE's metadata model already contains fields for referencing entries in common variable catalogues such as the *CF Conventions Standard Names* for climate-related quantities or the *World Bank's CETS* list of socio-economic indicators. We realize this should be extended by fields for referencing, e.g., the *CSDMS Standard Names*. We will check whether we missed any further important catalogues and add them if required.

A prominent framework that you should reach out to is the CSDMS BMI (basic model interface) idea. Your master component could implement that BMI/CMI and thus make all user-contributed

models also interoperable. We have, e.g., done this with the FABM (Framework for Adaptive Biogeochemistry) for ESMF interoperability. If you don't want to add a BMI (to CSDMS, OpenMI or ESMF, or other frameworks) please add a section outlining your plans to do so or your reservations against doing so.

This is a really very helpful hint, indeed we were sadly unaware of the existence of this initiative. In the next release of pycopancore we will aim at providing a generic wrapper component that allows wrapping external models that implement the BMI into copan:CORE model components, and will also think of how to implement the BMI ourselves in the base model component so that any copan:CORE model can run in a "passive" mode governed by an external coupler that calls its BMI. We will add a corresponding paragraph in the revised MS.

*p13 l 3ff The term "modular" is in your context the software modularity typically found in modern software architecture. This is *not* an emanating feature of copan:CORE. There is modularity beyond software modules in other frameworks and I would encourage you to rethink modularity in that broader sense.*

What we mean by modularity in the MS is (i) the division of the program *code* into packages representing "model components" that can be developed by independent "model component developers" and still use the same set of entity-types and attributes, "models" that can be composed from these components by "model composers", and "scripts" that "model end users" can use to perform specific studies, and (ii) the division of each entity-type's processes into contributions coming from different model components via multiple inheritance. We will try to identify further forms of modularity that copan:CORE does or should provide.

p16 l 4 A section on performance is missing (e.g. at end of section 3). Many thousands of cells, individuals or other entities might have to be simulated with this framework. What is your approach to ensuring that integrations of differential equations (exemplary for one of your process-types) is efficiently handled for large numbers of entities? Is there consideration for optimization (you already mention communication with MPI and JSON) for high-performance computing architectures? What tradeoffs to performance do you expect by using "slow" packages like sympy? Did you perform any scaling experiments?

We totally agree that performance is eventually a very important aspect for the production version of any software. With the current paper, the copan:CORE framework described therein, and its reference implementation in Python, pycopancore, our main aim is however a slightly different one than providing a performance-optimized production software. Such a performance-oriented production implementation of the copan:CORE framework, cppcopancore, is currently under development and its performance will be tested and documented thoroughly in a separate paper. For the revision – also in the interest of space – we will therefore limit our comments on performance to a sentence stating this and giving running times for the illustrative example.

Figures

Overall, the figures are of mixed quality and style. A more consistent layout, style, coloring and fonts across all figures would make the paper more pleasing to the eye and also more readable. Please spend some efforts towards this goal. Especially Figs 1 and 2 are very clear and could serve as a template.

We agree and will work on the figures to achieve a larger degree of stylistic consistency and aesthetics, taking Fig. 1 and 2 as prototypes.

fig 1 The white box could contain text, such as “none”

OK.

fig 2 For consistency with text, use “process type”, not “modelling approach”

Please see our response above.

fig 3 This entity–relationship diagram in UML style is only understandable to a small fraction of readers. Please explain the notation used in the diagram (for example by giving an example of the cell–person relationship). I do not at all understand the circular relationships for entities with themselves, especially for the SocialSystem entity. Please clarify. This figure does not need color, in fact, color distracts here.

OK.

fig 4 This “spaghetti” diagram is not helpful. Please create an entirely new graph. Rearrange the information, e.g., choose a UML style for consistency with fig 3. Avoid crossing lines, strange coloured shapes without obvious semantics, use typewriter font consistently for code parts. Make graphical markers (colors, line widths, boxes) easily accessible by adding a legend instead of explanation in caption.

OK, see also our response to Mr. Dermody.

fig 5 see comments for code figures later

OK.

fig 6 Change colours entirely to be consistent with figure 2 (CUL, MET, ENV). Don’t use background color. Change layout to something visually appealing; currently the table structure suggest as semantic for rows and columns that is not evident.

OK.

fig 7 Table layout conveys meaning, but could be highlighted (columns are scenario (is that what you call “runs” in the caption?, rows are taxa). Avoid mixing colour semantics with those of previous figures. Avoid mixing color semantics between panels: How to top and middle row colors align? If they do, don’t add two legends but use only one. Explain why for CUL/ENV there are only four quantities shown, but for MET there is an ensemble (each four) of three quantities shown. Upper left: where is the blue line (I guess hidden behind the grey one ...)? Find a way to display lines that are on top of each other without hiding any (also upper right figure). Possibly add events on time axis, especially for understanding middle right panel events with sudden transitions from fossils to biomass.

OK.

figs 5,8-10 Try improved syntax colouring and choose different font. Fixed width is important, but better use a smaller width. Consider light grey for comments, for example. A light (cream) background might help to set the code apart from the title, which is barely visible (and which uses inconsistent font with main text).
OK.

Technical comments

p7 l 10 There is no such thing as "sharp criteria". Criteria alone is sufficient.

Although we do not agree here, we will remove the "sharp" anyway.

p14 l14 The link to pycopancore (http://github.com/pik- 15 copan/pycopancore) does not work yet (so make sure it does work on publication day)

OK. The following link now works since May 2018: https://github.com/pik-copan/pycopancore . The online publication as open access code was delayed due to institutional legal checks that were pending.

p14 l29ff and Figure 5 Use a consistent form for presenting code, do not alternate between text and figure.

OK.

p16 l 4 Examplary => Exemplary

OK.

p16 l 9 "not intended to be a serious representation". A representation cannot be serious. I suggest "is intended to be a toy representation". BTW, what is the "real" world anyway :=)

OK.

p17 l3ff Avoid double parentheses throughout this paragraph and manuscript.

OK.