Manuscript prepared for Earth Syst. Dynam.
with version 2015/09/17 7.94 Copernicus papers of the LaTeX class copernicus.cls.
Date: 9 November 2015

Author's Response at Resubmission of Manuscript "Topology of sustainable management of dynamical systems with desirable states: from defining planetary boundaries to safe operating spaces in the Earth System"

Jobst Heitzig[1], Tim Kittel[1,2], Jonathan F. Donges[1,3], and Nora Molkenthin[4]

[1]Research Domains Transdisciplinary Concepts & Methods and Earth System Analysis, Potsdam Institute for Climate Impact Research, PO Box 60 12 13, 14412 Potsdam, Germany, EU
[2]Department of Physics, Humboldt University, Newtonstr. 15, 12489 Berlin, Germany, EU
[3]Stockholm Resilience Centre, Stockholm University, Kräftriket 2B, 114 19 Stockholm, Sweden, EU
[4]Department for Nonlinear Dynamics & and Network Dynamics Group, Max Planck Institute for Dynamics and Self-Organization, Bunsenstraße 10, 37073 Göttingen, Germany, EU

Correspondence to: Jobst Heitzig (heitzig@pik-potsdam.de)

Dear Editors,

we hereby resubmit the revised manuscript originally entitled "Topology of sustainable management in dynamical Earth system models with desirable states" after making the changes suggested by the four reviews.

This response is in part identical to our earlier "Final Author's Response" submitted during the initial review process after the first two reviews, with added answers to the additional two reviews.

Summary of changes

Since the main changes are a restructuring of the whole manuscript and the writing of some additional parts, we start here by giving an overview of the main changes made (instead of submitting a "latexdiff" file that would mainly be red simply because sections have been moved).

- Title and authors list have been extended.
- Introduction has been extended, new Sec. 1.1 and new Figs. 1 and 3 have been added.
- Main formal treatment has been assembled as the new Sec. 2, details have been united with the proofs in the Appendix.
- Main examples have been assembled in new Sec. 3 and put in more logical order, minor examples went to Appendix and the new Supplement.
- Discussion and Conclusions have been extended much.

Response to general comments

We thank the reviewers for their overall positive comments on the content of our manuscript and the detailed suggestions how the presentation might be improved. Our original idea was to introduce the needed mathematics not en bloc but spread-out throughout the paper, separated by examples that motivate and illustrate each of the pieces, somewhat like in typical lecture notes. From the reviews we realize that this strategy may not work as hoped and that a more strict division between verbal descriptions of concepts, their mathematical formalization, and illustrating examples may be better. For this reason, we restructured the manuscript as follows:

Before introducing the mathematical notation in what was Sec. 1.1 and has become Sec. 2 now, we added a new Sec. 1.1 with a concise verbal exposition of all our concepts, accompanied by what was the lower part of Fig. 3 (the summary of our concepts in the form of a decision tree) and has become Fig. 2 and by a much improved version of the upper part of the former Fig. 3 (as a motivation of our choice of metaphorical terms from nautics), the new Figs. 1 and 3, for whose design we were able to recruit the additional co-author Nora Molkenthin. We then continue in Sec. 2 with a more formal and detailed treatment of these concepts, choosing the lowest adequate level of mathematical complexity (and reserving the full mathematical treatment for what is now Appendix A1–A3). In contrast to what we suggested in our earlier "Final Response", to safe space we did not add another series of simple figures in the general spirit of what was Fig. 1 since we felt after the extended introduction these would not be necessary any longer.

Only after this, we present the illustrative examples in what became Sec. 3, stating each example's purpose more clearly than before, and moving what was Example 3 into Appendix B1. In addition to what we suggested in our earlier "Final Response", we also moved the former Example 6 into Appendix B2 so that Sec. 3 now only contains examples with a strong relationship to Earth System modelling (or classical physics in case of 3.5), and in an order more easily accessible for Earth System scientists, from a monostable carbon cycle model 3.1 via a multistable ecological model in 3.2 and a model of technological change towards renewable energy production in 3.3 to a conceptual co-evolutionary Earth System model in 3.4, closing with an example from classical mechanics included for its simplicity and wide range of phenomena, and a more theoretical example showing the relationship of our concepts to bifurcations similar to those occurring in tipping elements of the Earth System.

For a number of reasons we did not attempt at choosing different, even more realistic examples from Earth System Science but stuck to the above carefully selected set of conceptual low-dimensional models. First, as this is mainly a theoretical paper introducing new concepts and tools, the purpose of the examples is only to illustrate how the characterized state space regions and dilemmas may occur in various fields related to the Earth System. An actual application of the concepts to a realistic model of some subsystem of the Earth System is way beyond both the scope of the present paper and our abilities and should be the natural subject of future work. Second, in order to clearly illustrate the concepts graphically, the examples have to be two-dimensional. Third, in order to avoid unnecessary numerical complications and artefacts, and to give evidence for the fact that the discussed dilemmas may easily occur already in quite simple systems and are thus probably quite pervasive in more complex systems, their dynamics has to be as simple as possible. However, as some examples will appear unrealistic to experts in the respective field, in particular the former Examples 5 and 9, we took care to discuss them in more detail in the revised manuscript, also adding a Supplement with a detailed motivation for the ecological model.

The first referee judged that "the relevance of the framework to Earth system dynamics, and the relevance of the results for earth system governance, is also not sufficiently elaborated." Although we believe that we did elaborate quite much on the relevance for governance in the original introduction, the individual examples, and the discussion, we now decided in addition to what we suggested in our earlier "Final Response" to extend the respective discussion substantially in the new Sec. 4 with the help of another additional co-author, Jonathan Donges, who also contributed to the improvement of the Introduction and the additional text 1.1. Regarding our original bottom line message that "before performing some form of quantitative optimization, the sustainable management of the Earth system may require decisions of a more discrete type that come in the form of several dilemmata, e.g., choosing between eventual safety and uninterrupted desirability, or between uninterrupted safety and increasing flexibility," the second referee helpfully suggested to either make clearer "how tradeoffs between these criteria play out", and this is precisely what we now attempt in the extended discussion by comparing our concepts to the "sustainability paradigms" of Schellnhuber (1998) and by suggesting to use them to define a hierarchy of planetary boundaries instead of just one level. However, in view of the much more complex nature of the real-world Earth System, we don't feel experienced enough to make more reliable statements about the *actual* tradeoffs, which we believe may be a whole research agenda that may follow from our theoretical suggestions.

We are grateful for pointing out the literature on qualitative modeling of global change patterns and related issues such as the tolerable windows and guardrails concept and were happy to add the corresponding references.

Response to specific comments and questions from the "technical corrections" section of reviews 1 and 2

(Page and line numbers refer to the original version)

Page 442, line 15: A reviewer asked, why is the reachability between two sets limited to both being "arbitrarily close"? Wouldn't reachability also make sense for any pair of sets, whatever their distance? This seems to be a misunderstanding. We define a notion of stable reachability by giving certain mathematical conditions under which we say that "a set Y is stably reachable from a state x through some other set A". Y and x do not have to be arbitrarily close (but Y and A do have to, since otherwise one could not reach Y without leaving A). For technical reasons, we formulate the condition used in the definition of "stable reachability" in terms of another, auxiliary notion, that of "forecourt", which allows us to elegantly define "stable reachability" like this: A set of states Y is stably reachable from a state x iff (meaning "if and only if") x is in some forecourt C of Y. Now the words "arbitrarily close" which apparently confused the referee only occur in the definition of "forecourt": A forecourt of a set Y is another set C from whose every point one can approach Y arbitrarily closely without leaving C by suitable management.

As a consequence, the forecourt C must of course be arbitrarily close to Y in the sense that it intersects every topological neighbourhood (in particular, every ε-neighbourhood) of Y. We could alternatively have avoided the notion of arbitrarily close approach and instead have demanded the more restrictive condition that one can navigate properly *into* Y from every point in C without leaving C. In that case, the forecourt C would even have had to intersect the target set Y. But then all asymptotically stable fixed points which one cannot ever reach exactly in finite time but can only converge to in infinite time (and hence can only approach arbitrarily closely in finite time) would not have been counted as "stably reachable". However, such asymptotically stable fixed points occur in many models, and in practice (and in view of typically unavoidable small perturbations) the possibility of an arbitrarily close approach is all that will ever be relevant for real-world management. For this reason we chose the more complex definition involving the condition of arbitrary close approach instead of the simpler but more restrictive one.

Page 446: Example 3 was only given for completeness as a concise summarizing example for the more mathematically inclined reader whose main value is that it shows that all the introduced regions can occur simultaneously even in a simple low-dimensional example with a high degree of symmetry. It was however not considered important enough to spent space for an accompanying figure. We decided to move this example into the appendix in the revised version of the manuscript.

Page 449: Example 5 on alternative plant types and multistability is meant to show with a conceptual model how a lake dilemma may occur in a simple multistable system. Although it is known that many plants modify the soil in ways that benefit their own growth, e.g. via influencing microbial communities and biogeochemical cycling (e.g., Kourtev et al., Ecology 2002; Read et al., New Phytologist 2003) and empirical evidence exists that this has effects on interspecies plant competition (e.g., Poon, Master's Thesis University of Guelph, 2011), we know of no formal model that would allow to study the resulting feedbacks between two plants and is simple enough for the purpose of illustrating our theory in an adequate amount of space. The best existing candidate models seem to be the four-dimensional model of a two-species plant-soil-feedback by Bever et al., New Phytologist, 2003 (see also Kulmatiski et al., J. Ecol. 2011) and the spatially resolved model of an invading plant by Levine et al., Ecology 2006, which however does not model other species explicitly. For this reason, we chose to design a conceptual model of two fictitious plant types each of which grows according to the well-established logistic growth dynamics leading to an initially exponential growth that is dampened by intraspecies competition. In order to keep the state space dimension at only two dimensions so that state space diagrams can be plotted, we refrained from modelling the soil characteristics via dynamic variables as in the other models, and instead represented the soil modification effect by simply assuming that the two species' undampened growth rates are proportional to some carrying capacities K_1, K_2 that the current soil composition implies for the two species, and that K_1, K_2 depend directly on the existing two populations x_1, x_2 in some simple way. In order to study the effect of soil modification alone, we did not include other interspecies interactions such as direct interspecies competition for resources. Levine et al., Ecology 2006, also assume dampened growth with a basic rate that depends on the existing population, but they only focus on a single species and assume a fixed carrying capacity, which we find somewhat implausible in view of the empirical evidence presented in Poon, Master's Thesis University of Guelph, 2011. Because we wanted to produce a conceptual model that illustrates the topological landscape in a multistable system, we needed to make sure the actual functional form we chose for K_1, K_2 produces a multistable system. This was achieved by assuming that the effect of the two populations x_1, x_2 on the two carrying capacities K_1, K_2 is nonlinear in the sense that the marginal soil improvement by plants of the same species is declining with higher populations while the marginal effect of plants of the other species is increasing with their population. We are not claiming that this is so in real-world plant-soil-feedback systems, but believe that the alternative assumption of a linear relationship seems unlikely. We then chose a very simple formula for K_1, K_2 that has these properties. In the revised paper, we have made these choices more explicit and added the above discussion and references as a Supplement.

Page 455-461: We believe we followed the suggestion to "clarify better in words what some of the term mean (transitional, rapid, channel, fairway, finer partition)."

Page 464, line 5: "What situations in the Earth system might correspond to lakes, glades, etc.?" This is a hard question probably requiring specialists from several fields of Earth System Science, so we were only able to speculate here. We believe we now did so in a transparent way in the new Sec. 1.1.

Further technical corrections not yet discussed above

We performed additional spell-checks and were happy to change the plural of the greek word "dilemma" from its greek plural "dilemmata" to the English "dilemmas" as suggested. The mentioned German word "beim" is part of the cited author's surname and thus correct. Regarding figure sizes, we had designed them originally already to work well with the eventual ESD two-column layout instead of with the ESDD landscape layout, which explains the somewhat suboptimal appearance in the discussion paper, for which we apologize.

In the revision, we took care that all relevant details are legible.

One referee pointed out a possible misunderstanding of equations of the type $S = (a,b)$ in which a set S is equated to an open interval (a,b) of real numbers since the notation (a,b) could as well refer to a point in 2D space. We took care to make this unambiguous by using the alternative notation $]a,b[$; the use of the equal sign to denote equality between sets (as between any other mathematical objects) however is so standard and needed so often in the paper that we could hardly imagine how to avoid it.

Page 450, last two paragraphs: We thank the anonymous reviewer for the comments regarding this discussion of the two stable fixed points, the shelter, and the lake. We reformulated them in a more accessible way to avoid the pointed-out ambiguities.

Page 451, line 25: We introduced $g(r)$ as the upper solution to the equation $F(g(r)) - 1 = 0$, and elaborated on this a little more in the revision.

Page 453, line 4: Following the reviewer's suggestion, the names and a short explanation for all the individual parameters have been included. With the normal serif font ESD uses for their final papers, the greek lowercase gamma and the letter y should now be easily distinguishable, so we decided not to change the symbol.

Page 453 line 10: As it was correctly pointed out, values of $y < 0$ are unphysical, so the correct bound on y should read $\max(0, -(\ell + \delta)/\phi\gamma_0)$ instead of just $-(\ell + \delta)/\phi\gamma_0$, since the latter may be negative depending on the relative sizes of the negative quantity ℓ and the positive δ. This has been corrected.

Page 457: As mentioned, Example 9 was designed by us only to illustrate the relationship of reachability and bifurcations. It is not meant to represent any real-world example. However, it contains an individual and a paired saddle-node bifurcation similar to those that may occur in bistable Earth system components such as the Atlantic Meridional Overturning Circulation (AMOC) or other tipping elements (Schellnhuber, 2009), as mentioned. We added a few references to make this even more clear.

Page 460 (!), line 18: We now use a different letter than D for docks to avoid the pointed-out collision in notation.

Page 479, figure 1: We moved this figure into the Appendix, where it will be easily understandable after reading the article.

Page 481, figure 3: Upper part: This figure was meant to depict all introduced regions in an illustration of a state space that roughly resembles a flow from an "upstream" via a "downstream" region into an "abyss" etc. Obviously, it was much too abstract to really make this intended analogy apparent. Since we still believe that such an illustration can help motivating our choice of metaphorical terms, we replaced it by a much more explicit drawn illustration of a water system involving a river running from a mountainous region into an ocean with a trench, what is now Fig. 1. The lower part is now Fig. 2.

Page 484, figure 6: This example has been moved into the Appendix, and we have made clearer than before that the four subplots are for four different archetypical systems which represent archetypical situations of bifurcations. We have also added more labels and vertical arrows to indicate the default dynamics towards the stable fixed points.

Page 485, figure 7: We thank the anonymous reviewer for the comment concerning γ_0 and γ_1. These refer to the values of the parameter γ in the default, unmanaged dynamics and the managed dynamics, respectively. We clarified this in the text. The other parameters refer directly to their mentioning on page 453, line 4.

Page 487, figure 9: By "extreme admissible management trajectories", we mean in this context those admissible management trajectories that correspond to the largest-possible leftward or rightward motion from their initial point. We chose a less ambiguous phrasing in the revision.

Responses to the third and fourth review

These reviews were added after the initial end of the discussion and our "Final Author's Response" and were hence not yet responded to. Referees' text is in quotes.

Response to referee 3

"Firstly, the introductory statements on economics could be misunderstood as economic theory being systemically ad odds with the insights presented. While this may hold for standard economics being based on expected utility (maximization) theory, in particular in the climate context a major fraction of the community has produced the majority of results in a lexicographic manner (see e.g. IPCC AR5 WGIII Ch6)."

We appreciate this comment especially since it was not our intention at all to make statements about economics or claim that economic theory was at odds with our insights. Rather, we believe that the qualitative assessment our concepts enable might work very well in combination with established tools particularly of the integrated assessment community. In order to avoid this misunderstanding, we have now added another paragraph discussing this relationship and some references to the introduction.

"IPCC AR5 WGIII Ch2 in turn points to the price one has to pay when deviating from expected utility maximization. While lexicographic preferences have a certain appeal for system-theory trained researchers, they deliver very unattractive features when being applied in the context of dynamic decision-making. Hence the authors should at least consider the following 4-phase type of analysis: (i) topological analysis as presented, (ii) putting up a welfare measure that is sensitive to this rich topology, (iii) optimization, (iv) clarifying to what extent this could still be interpreted as a market equilibrium, and if not, inclusion of suitable policy instruments."

We elaborated on this a bit further now in the discussion.

"In particular the authors seem to be ignorant of the work on economics of tipping points as being performed in the groups of Klaus Keller (Penn State U), Tim Lenton (U Exeter), Martin Quaas (U Kiel)."

We are aware of these groups' important work but still feel it falls well within the mainly quantitative analyses we mentioned since their main point seems to be that economic models need to take tipping points into account, which is obviously true but a very different question from how the resulting models should be analysed.

"Secondly, the ms delivers a great service to the community in making rather abstract content accessible to a wider audience. However it is not always so clear (i) what is new in terms of content / theorems, (ii) what is merely a didactic service, a transfer of knowledge from one community to another. In particular for me it was not so clear in what it differs from previous work by Aubin and his group, or Schellnhuber."

We apologize for this ambiguity. In the new introduction, discussion, and supplement we now make this much clearer.

"1. The intro should also cite the tolerable windows approach (Petschel – Held et al., Climatic Change, 1999; Bruckner Zickfeld, 2008)."

We added respective references.

"2. 1.1 tends to introduce terms on a non-technical level first. However I find this confusing. Keep to the standard scheme: definition first (that should be illustrated, indeed), then usage of that term."

Since this issue was also raised in somewhat different ways be other reviewers, we had to make a hard choice here. After many presentations and discussions with researchers from different subfields, we were convinced that this "mathematician's" style would not work so well with the intended audience without a verbal introduction of our concepts first. Hence while separating definition and usage as suggested, we deviated slightly from the above proposed scheme by first introducing all concepts informally in the new Sec. 1.1 before introducing the suitable formalism in Sec. 2 and applying it in Sec. 3.

"3. Figures should be added to 1.1 and following subsections." "8. Figure 3 is cited before Figure 2."

We rearranged the figures properly as outlined above.

"5. P440: Has 'safe' been defined before usage?"

That very sentence is meant as our working (verbal) definition of "safe". We added italics to make this clearer.

"6. P442: We need a figure on C. The whole technical concept at the center of that page should be shifted to the appendix, and the essence should be illustrated."

We agree that this technicality was better placed in the Appendix but did not add yet another figure to save space and since forecourts are not used anywhere else than for the definition of stable reachability.

"7. While Fig1 is extremely helpful in general, I found a couple of aspects confusing. Hereby I assumed there was a water flow from the left to the right:"

This is a misunderstanding, there is no water flowing in this example, the blue color is only used to be consistent with the other plots when indicating default and alternative trajectories. The default dynamics is downward as indicated by the two pale-blue arrows.

"a. What is the difference between solid and dashed thick lines? Can we have a fixed point for the latter?"

We apologize for this graphical conversion error – the dashed thick lines should have been solid as well.

"b. Why is 'sunny downstream' downstream? How can we return along a thick line to the manageable region?"

It is downstream since from there one cannot cross the point a to reach the shelter (if one could, one would be in the upstream) but *can* reach the manageable region (hence one is not in the eddies or abyss).

"c. Why is a thick line before 'a' a subset of the shelter? If the flow is always downstream how can the boat stay in front of 'a' w/o management? Is there a fixed point?"

Since default movement is downwards in the picture, there is indeed a fixed point at the meeting point of the thick and thin lines just left of point a, and since this fixed point is sunny, its basin (the region between the two adjacent summits) is a shelter as indicated.

"d. Are the colors at the boat's arrows messed up?"

We believe not: pale blue indicates default movement downward (which is in this case leftward), dark blue dashed indicates alternative movement upwards (which is in this case to the right).

"e. Why are the glades as illustrated here not a subset of the shelter?"

Since the default dynamics lead away from the shelter, only by management one may reach the shelter.

"9. I think the logic in the § before Eq6 is messed up. Should it not read somehow like: 'Now we turn to the region from where one cannot avoid ending up in the trenches. We define the abysses Y (uppercase greek letter Upsilon) as the closure of this region, in order to robustify it against infinitesimal perturbations. ...' Then , of course, minus trenches. My point is: 'if one has to fear' comes too early."

This is of course correct, we removed the first occurence of 'if one has to fear' since it was redundant.

"10. Eq7: Clearly indicate that you are moving beyond Fig1."

In the new Fig. 1, eddies are contained.

"11. Example 2: clearly say whether a(t) is management or system."

We do so now (it's management as implied by the equivalent term "alternative trajectory").

"12. Motivate the structure of Ex3 better."

This example was less important than the others and has been moved into the Appendix.

"13. § below Eq16: Why is the last term in U^- a $(+)$? Why do we always end up in $(+)$ w/o management?"

The definition of upstream U is that one can navigate to the shelters, which are in this case equal to the region $(+)$. If no management is possible, this means all trajectories starting in U *must* end in $(+)$, otherwise their starting points would not belong to U but to either D, E, Υ, or Θ by definition.

"14. Ex4: Where are anthropogenic emissions?"

Since $c_a = 1 - c_m - c_t$, the extraction term $-\alpha c_t$ from $\dot{c}_t$ occurs as an emissions term in $\dot{c}_a = 1 - \dot{c}_m - \dot{c}_t$.

"Also the last § [of Ex4] should be expanded in view of the audience of ESD."

We extended the discussion of this example as suggested.

Response to referee 4

"Many of the figures I find only partly understandable. Particularly Fig. 1 I find too Cdetailed and explained insufficiently, even though it is of great importance in guiding the reader through the manuscript."

The new Fig. 1 is much improved and discussed in more detail.

"Overall, I find the manuscript almost trivial, and I think there is too much emphasis on the math."

While we do not agree on the level of triviality, we have moved much of the formulas to the Appendix since the emphasis is actually on the qualitative differences between the identified regions and the resulting dilemmas.

"A shorter paper with a focus on the basic idea with less, but more un- derstandable examples would in my opinion be fully sufficient. Especially the repeated focus on open sets may be mathematically more precise, but it is unlikely to play a role in applications, so that the concepts could be described in a simpler way."

We did so in the added Sec. 1.1. Regarding the "openness" question, we feel that in contrast to viability theory, which uses closed sets and needs some more efforts for proving the respective theorems, our usage of open sets instead simplifies matters. Whether the related question of stability plays a role in applications, we cannot answer at this point, but one can easily imagine systems with unstable equilibria, and since those should not be classified as "safe", we need the stability and thus the openness requirement.

"I found the distinction among the dilemmata the most interesting aspect of the paper, while the remaining parts did not provide me with insights. I could imagine that a focus on the dilemmata could strengthen the paper. I found the reference to Earth system models in the title confusing."

We are very grateful for these comments and extended both the discussion of the dilemmas and changed the title for clarification.