

## Reply to anonymous referee 2

*We would like to thank our referee for the very helpful comments which will improve our manuscript! If not stated otherwise, our references to pages, figures, etc. are based on the submitted, not the revised manuscript in order to match the references of the reviewers.*

*We quote the referee's comments one by one, each followed by our reply in italic font.*

### General comments:

1. My major concern with the present paper is the high complexity of the method. In parts it is difficult to follow, and it would definitely benefit the paper's readability. I would propose to have an appendix of a clear step by step example. This would improve the understanding greatly.

*We have restructured Section 4 accordingly. See also our reply to referee 1 which reads: "Sect. 4 is now structured as follows: As before, we introduce system 3 in Sect. 4.1. In Sect. 4.2 we introduce the general framework of hotspot detection which is the same for all cases. In Sect. 4.3 we then introduce the different versions of the method and discuss their advantages and limitations. Both, Sect. 4.2 and 4.3 refer to a simple example which is given in Table 3 and a new figure (now Fig. 7). All settings are now listed together in Table 4 (previously Table 4 and 5) and the different choices are numbered to be referred to in a simpler and more precise way in the text. The justification with regard to the choice of EOFs is also included in Sect. 4.3. We also note that our example in Table 3 has changed as it is more instructive to present three different types of elements."*

*We have also included a step-by-step example. As it is referenced quite often in the text and illustrates the different steps, we prefer to present it within the main paper and not in an appendix.*

2. what would happen if elements are not removed?

*See point 17.*

3. Given this, I am wondering on how better the method performs in case 3 when compared to measuring signals directly from the different elements. It seems that as different elements have different stability properties, and since the time series are available one could compare the signals directly without going the proposed complicated protocol.

*It is an essential point in our manuscript that the use of local EWS generally does not work. See Sect. 3.2 and point 13 for an explanation.*

4. In the same way, isn't the mean of each element already telling something? Is there not perhaps more information in the combination of mean and ews derived directly from the elements?

*As mentioned before, local EWS do not work in a general setting. We also do not consider the mean state as a useful and generic EWS as the absolute value of a certain equilibrium is unrelated to its stability. For any system, the mean state can easily be shifted by introducing a coordinate translation. Similarly, the means at different elements do not allow inferences on the spatial interactions. For*

*example, in a 2-element-system with element 1 as the hotspot, state  $V_1$  can be higher or lower than  $V_2$ , depending on how parameters  $P_{0i}$  are chosen. Using the mean as an indicator hence only works if the behaviour of a particular system is already known.*

5. The authors show that if the signal becomes weak in case 2 but it is the same in all elements. I may be missing something but it would add to their argument if one compares signals directly to infer the different hotspots than compared to their method.

*Again, see Sect. 3.2 and point 13.*

6. The results are presented for two cases of noise. Even if the difference in the noise source is an important issue, how possible it is not pertinent to the specific model? In practice it is difficult to discriminate additive vs multiplicative, usually both are present. In that case what would be the best approach? And how the method will perform? In line with my previous observation, to simplify matters, wouldn't be better to present one method? And move the other in the appendix? Or even better, show the performance of the method when both sources of noise are present?

*We do not include any experiment with both noise types at the same time. In our case, the multiplicative noise would dominate the results and they would look similar to the pure multiplicative noise case. Also, our aim is to discuss the effect of both individually. We therefore do not think a mixed experiment to be that important.*

*However, we now discuss this issue by adding in our new Sect. 4.3:*

*“In a more general case, additive and multiplicative noise may occur at the same time. In our system 3, the multiplicative noise would dominate the results if noise levels leading to similar variance in  $V$  were chosen. However, it is not a priori clear what would happen in other systems whose properties are not well-known. Under such conditions the generic approach using cross- and autocorrelations with SDI and ERI would be the safest option in the light of our results.”*

Specific comments:

7. p646, line 7: spatial EWS are one way of overcoming issues with estimating EWS in timeseries. not necessarily for alleviating the issue of inadequate sampling or incorrect (with relationship to the right timescales)

*What we meant here is the following: Spatial EWS use each time step as a sample to infer the stability, while temporal EWS need a window of many subsequent time steps. As forcing changes over time in transient cases, the latter thus involves information on previous states of the system. It is often argued that therefore, spatial EWS can (under the right conditions) provide a more precise estimate of the current stability. We agree that this argument is unrelated to the problem of filtering out the right time scales. Apparently our wording gives this impression. We have therefore reformulated the paragraph accordingly.*

8. p 648 l 20: what is the meaning of the timescale tau?

*We have added the following explanation:*

*“The timescale tau describes how fast vegetation cover can establish in previously unvegetated areas (or die back in vegetated areas). Following Liu (2006) we fix tau to 5 yr which is meant to represent the dynamics of grass in arid subtropical ecosystems.”*

9. p650 l 20 I think it would interesting to show the P value as B is changing. In that way one can see that the P value of grid cell 2 slowly changes and then jumps to a lower value after the collapse of grid cell 1. Now in figure 2 parameter B represents the distance to the transition for element 1 and not of the whole system. Showing the actual Pi values for both elements would have made clearer the loss of stability in the system and thus the relationship to the EWS.

*We cannot completely follow this argument. We agree that B represents the distance to the transition for element 1, but the latter is identical to the transition point of the whole system because element 2 experiences an induced tipping (and has no tipping point on its own). The P vs. B curves (see our modified figure below) look similar to the  $V^*$ -curves in the original Fig. 2a, with the difference that the decrease in element 1 relative to element 2 is steeper due to the different parameters  $s_i$ . However, this fact alone does not tell anything on the causality of the tipping which is determined by matrix  $k$ . We could of course add the P-curves to Fig. 2, however as all the analysis is based on  $V$  rather than  $P$  this may confuse the reader.*

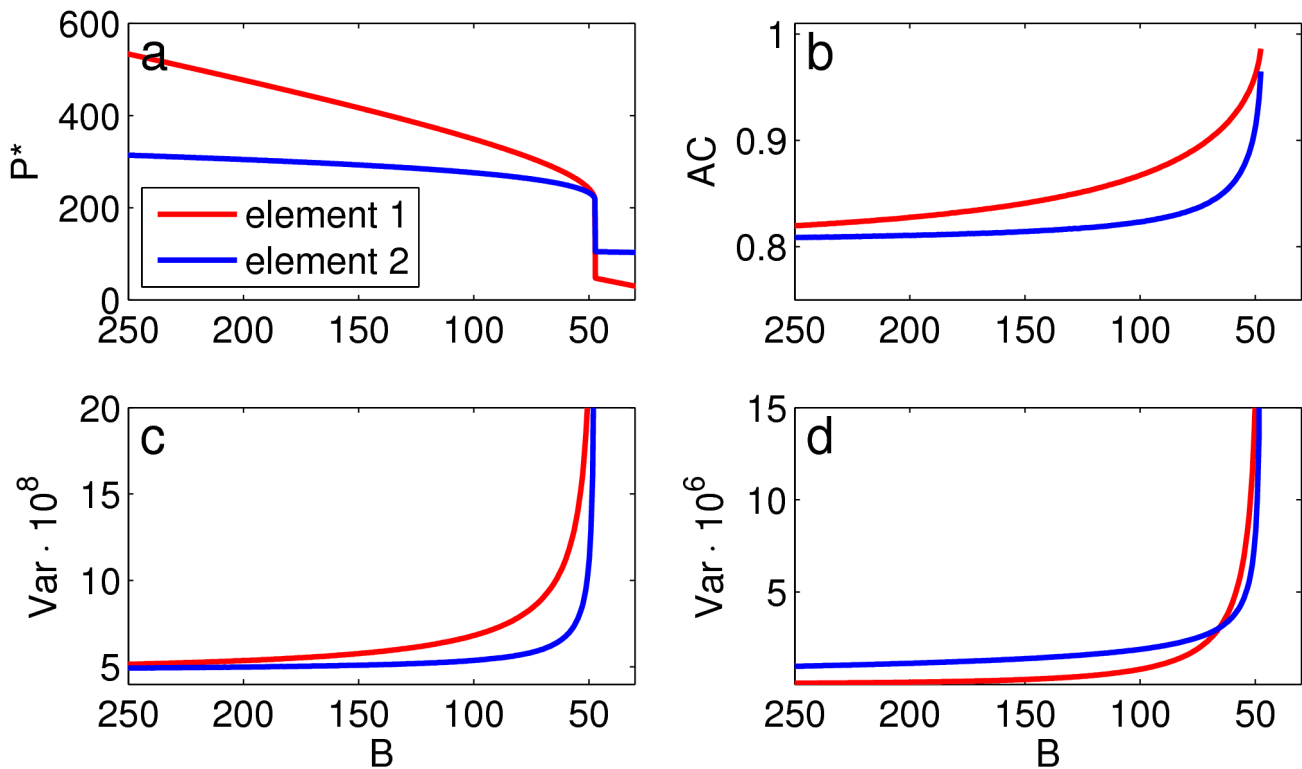


Fig. 1: As Fig. 2 from our manuscript, but with  $P^*$  instead of  $V^*$  in (a).

10. p651 and figure 2: I think you see the signal for both elements one more than another as in one case the system is going close to the bifurcation itself, while in gridcell 2 only partly and then it jumps to over the transition. I m not sure that the increase in variance wouldnt be seen only with the additive noise when there is no feedback. It would practically be invisible but theoretically the system would be approaching the bifurcation so the EWS should change.

*We agree that some increase in variance does exist for element 2.*

*What we mean in line 9-11, page 651 is the following:*

*Consider a 1-element system. If there is no feedback ( $k=0$ ) and only additive noise there cannot be any EWS and there is no bifurcation.  $V^*$  then simply follows the green  $V^*$ -curve as in Fig. 1 of our manuscript.*

*To clarify this we change the according paragraph as follows:*

*“In case of a single isolated element without any P-V-feedback ( $k=0$ ) there would still be an increase in variance in the multiplicative noise case, but not in the additive noise case. In our system 1, the slowing down at element 1 also affects element 2 due to the interaction term. This is the reason for the rise of the blue curve in Fig. 2c.”*

11. Perhaps concentrate on one source of noise (multiplicative that you suggest that has more sense, or combine both) and makes the paper more simple?

*We agree that this would make the paper simpler. On the other hand, both types of noise can occur in models. Our aim is to present the hotspot detection method in its general version (projecting everything on EOF1 of the last time slice), but at the same time to point out the advantage (much shorter time series needed) and the disadvantage (system-dependency) of possible modifications. As our modified approach works due to the effect of multiplicative noise in system 3 we need to present both types of noise in the manuscript (but independently to disentangle the effects).*

*In order to make this motivation clearer we have altered the structure of Sect. 4 as explained above.*

12. p651 l25: It should be clearer written why the elements are not bistable anymore. I guess the authors mean that would the system not be connected, the individual elements would not be able to shift under the current parameterization?

*Correct. We have changed this paragraph as follows:*

*“As local feedbacks (determined by  $k_{ii}$ ) are weak, no single element would be bistable anymore if all other elements were fixed.”*

13. p652 Why does the signal become less strong?

*To explain this phenomenon more clearly we have added the following paragraph:*

*“The critical mode implies the direction in phase space in which the bifurcation occurs. Slowing down particularly occurs for this mode and can be revealed by*

*the appropriate projection. In contrast, other modes of the system's variability are not necessarily influenced by slowing down as the changes of the stability landscape in other directions (characterised by changes of the according eigenvalues) are unrelated to the bifurcation. Hence, EWS in projections on other modes cannot be expected. The analysis of local EWS at individual elements would generally contain information on these other modes of variability and would therefore be a futile strategy. It has to be concluded that if the critical mode of the transition is not known beforehand, the tipping can be unpredictable even in cases of very long time series."*

14. why not symmetric the effect of the impacts of one element to another table 2? wouldnt that be more realistic for this model?

*Our examples (system 1 and system 3) address the problem of asymmetry on purpose. This is an essential point of our manuscript and we argue that this is the more realistic type of setting. For example, the impact of a land cover change in some area 1 on the climate in some area 2 is of course not identical to the impact of a land cover change in area 2 on the climate in area 1.*

15. p655 the number of parts is given by  $N/n_{\max}$ . in the example  $N=25$  and  $n_{\max}=3$ . Should not that be 6? This part is a bit unclear? What is an area? In the algorithm it is supposed the small number of  $N_p$  to 2, but in the table there are areas with only  $N_p=1$  (like 13, 18, 23).

*$25/3$  is 8.333..., of which the ceiling function (smallest integer larger than 8.333...) is 9. Hence, the number of parts is 9.*

*One should not confuse the size of a part ( $N_p$ ) with the number of elements selected from this part (between 1 and  $N_p$ ). Apparently, we should improve this explanation.*

*To make our nomenclature more clear we reformulate paragraph B1:*

*"For a given part of the system with  $N_p$  elements, we select a subset of  $n$  elements from these  $N_p$  elements. We refer to this subset as an area. Hence, there are three levels of selected elements where each set is a subset of the previous one: The number of elements in the complete system  $N$  (here: 25), the number of elements in a part of the system  $N_p$ , and the number of elements in an area of this part  $n$ . Example: We choose elements 19, 20 and 25 as a part. Hence,  $N_p=3$ , and  $n$  can be 1 (3 possible combinations), 2 (3 possible combinations) or 3 (1 possible combination)."*

16. p656: part D: when an element is thrown out of the analysis, the EOF is calculated in the other parts but without the element that was eliminated?

*Basically yes. To clarify this we add to paragraph C:*

*"The removal of an element means that it is not considered to be part of the system anymore and is not used from that point on. In this sense, the total system size  $N$  successively decreases and with it the number of parts is also reduced automatically."*

and on p. 656 (after paragraph D) we add:

*“After each calculation of the signals in all possible areas of all current parts and the potential removal of elements the remaining system is subdivided anew, starting from step A (large loop in Fig. 6).”*

17. p 656 l 19 How important is the elimination of elements for the results? If eliminating elements as the authors suggest emphasize the contribution of the hotspots to their identification, isn't this creating bias to the results? why are estimates of all elements present in figure 7 if elements are eliminated?

*Creating this “bias” is actually the reason why the method works at all. The “bias” is therefore not an error in the detection but rather the very thing that we want to detect.*

*All elements of the system are sampled, but those which are removed early are less well sampled than the others. We explain this in our reformulated paragraph as follows:*

*“The procedure serves as a sieve in order to filter out the important elements with a sufficiently small number of calculations. Without the removal of elements, the number of possible combinations would be too large to achieve a robust hotspot detection within a feasible amount of time. As the results depend on the random distribution of elements to different parts, they will be very similar but not completely identical when the analysis is repeated. The hotspot of slowing down can be identified if the time series are long enough (or if enough realisations are available), because the remaining elements at the end of the analysis tend to contribute most to slowing down.*

*To obtain more quantitative results, all signals calculated during the procedure can be collected in a sorted list for further analysis. Elements belonging to the hotspot tend to be part of the areas with the strongest signals and are on top of the list (Fig. 10). However, elements that have been removed early during the analysis are not well sampled. The method therefore only provides information on the nature of the hotspot, but less on the stability properties of the rest of the system.”*

18. p657: why is the covariance matrix better to use than the correlation for the the multiplicative case? is it perhaps so only for this model that multiplicative noise has higher sensitivity to the model outcome? wouldn't be better to estimate EOF on both additive noise as well?

*See p. 658 l. 1-17. In our new Sect. 4.3 we address this more explicitly with more specific reference to the question of the EOF:*

*“EOFs can be calculated as an eigenvector of the system's covariance matrix or alternatively its correlation matrix. If based on the covariance matrix, elements with large variance will be emphasised. Whether this improves the performance of a hotspot detection generally depends on the system under analysis. In case of system 3 with multiplicative noise,*

variance is enhanced particularly at the hotspot when  $B$  approaches the bifurcation point. Therefore, SD2 with ER2 yield the most significant results when using covariance-based EOFs. “

And later on in this section we write:

“In system 3 with multiplicative noise, SD2 and ER2 in combination with covariance-based EOFs are of particular advantage. Fig. 11 shows the signal list for the multiplicative noise case when using time series of length 10.000 yrs and options as set 21 in Table 4. While for the additive noise case the AC's trajectories at the hotspot (red area) and the complete area always look alike (not shown), they differ substantially in the multiplicative case: At the hotspot the signal starts to emerge early, even when projecting on a leading EOF far from the Tipping Point (red curves in Fig. 9). This is not the case for the other areas because the variability of the system differs substantially from the critical mode. As variances at the hotspot are very small, the EOF pronounces other elements than the hotspot and slowing down will thus not be observed in the projection. Close to the Tipping Point, the variance at the hotspot increases not only due to slowing down, but also due to the multiplicative noise which enhances variance as vegetation cover decreases. Therefore, the relative increase in variance is particularly large at the hotspot. Close to the Tipping Point, the system's variability becomes dominated by the critical mode and slowing down can be seen in the complete as well as the hotspot area. By using SD2, ER2 and covariance-based EOFs we use this property of the system to better distinguish the elements from each other. As a result, the hotspot can be detected much easier than in the additive noise case. Using time series of 10000 yrs each, the hotspot is clearly visible in the signal list for  $n_{\text{max}}=5$  (Fig. 11). Hence, an even more robust hotspot detection can be achieved from time series ten times shorter than in the additive noise case.”

19. p658: would that mean that one would need at least 10000 points for the method to work?

*This is correct for our example system. However, other systems may be simpler and could yield more significant results. Part 2 of our manuscript provides an example for this.*

20. p658 l22: Which are the different conditions?

*They are listed in Tables 4 and 5.*

*We now write:*

*“For a quantitative comparison of the algorithm's performance under different conditions (parameter settings, choice of detection method and time series properties), we perform 500 Monte Carlo experiments for each condition, using lag-1-autocorrelation (Table 4) and relative variance increase (Table 5) as EWS.”*

21. in all figures with transitions: does the elements in all systems shift at the same time?

These are equilibrium values plotted and it is interesting to see if in the real timeseries

there is a lag. Perhaps it should be made clear in the figure captions.

*In addition to our explanation on p. 652, l.23-24 we have added the following remark on p. 650:  
“The curves in all our figures are derived from stationary time series. However, if B was very slowly reduced during an experiment, the transient time series of the collapse would follow the equilibrium curves like in Fig. 2a very closely because the noise level is small and because the timescales of both elements are identical and small compared to the parameter change. Therefore, it would not be possible to infer the causality of a transition from the timing of the collapses at different elements.”  
To prevent a repetition of this paragraph we refrain from adding it to several figure captions.*

Technical corrections

22. p650 l2: units of sd of noise difficult to understand the meaning of k (in units?)

*We have changed this and now comment:*

*“For simplicity, we provide  $P_d$ ,  $k$  and  $\sigma_P$  without units, although the value of  $P$  represents mm/yr.”*

23. fig 3 axis x labels missing.

*Our original Fig. 3 was cut by the typesetting team to fit the discussion format. As the final paper will have a different format we choose to wait for the final typesetting.*

24. note that all elements have the same measured indicators

*Good point. We have added this note to the figure caption.*

25. fig 10: should better be occurrence than frequency?

*We have changed this.*