

Earth Syst. Dynam. Discuss., referee comment RC1
<https://doi.org/10.5194/esd-2022-41-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on esd-2022-41

Anonymous Referee #1

Referee comment on "Reliability of resilience estimation based on multi-instrument time series" by Taylor Smith et al., Earth Syst. Dynam. Discuss.,
<https://doi.org/10.5194/esd-2022-41-RC1>, 2022

This manuscript addressed an interesting topic by investigating whether measurement noises can impact the inference of resilience using remote sensing data. They used a simulation approach to investigate how signal-to-noise ratio (SNR) influences the calculations of two indicators of resilience, namely lag-1 autocorrelation (AR1) and variance. Their results have implications for assessing the possible impact of measurement errors in observational data. Overall I found this study well designed and conducted. I have a few comments that may help improve the paper.

(1) the study generated simulated time series by combining a background time series with a number of random noises. I was wondering if this characterized the realistic errors introduced by changing instruments. I was not an expert in remote sensing, but I thought in some cases the change of instrument might induce sudden increase/decrease in the time series (rather than a small noise term). Such changes can have major impacts on the calculation of resilience indicators, keeping in mind that such indicators were used to detect 'sudden changes' in the time series, whether they were due to measurement errors or underlying processes?

(2) AR1 and variance are two important indicators of resilience, or early warning signals (EWS) for catastrophic changes, but there are more. Moreover, researchers had been developing composite EWS by combining different metrics. Given that measurement errors may influence AR1 and variance differently or in opposite directions, I was wondering if a composite EWS would be more robust to measurement errors.

(3) the authors discussed about the difference between the average of variance from individual time series and AR1 of the aggregate time series, particularly their different behaviors in the presence of measurement errors. Similarly, the reference Feng et al. (2021) found different temporal trends of these two metrics. However, these two metrics represent different properties (i.e., local- vs. larger-scale resilience) and they did not necessarily exhibit different patterns, even if there was no measurement error. The problem is, the local-scale variance did not add up to give the larger-scale variance, but modulated by the synchrony between local grids. I attached a theoretical paper illustrating this:

Wang, S. & Loreau, M. Ecosystem stability in space: α , β and γ variability. *Ecol. Lett.* 17, 891–901 (2014).

(4) while the manuscript was overall well written, I had to say that I was confused by the different metrics involved in the figures, which seemed to be quite related but differ in important ways. For instance, the authors calculate resilience indicators using several approaches, e.g., deriving the numbers for an individual time series, first aggregating the time series and then calculating AR1 and variance, or first calculating AR1 and variance and then averaging them. They also calculate correlation between AR1 and variance at different levels of complexity. I would suggest to add a table to clearly define all key metrics in the figures, with explanations what a positive/negative or higher/lower value mean.

Specific comments:

L52: What does 'synthetic series' mean? I think it is simply a simulated time series.

L79: How was this 'aggregating' implemented?

Figure 2: Not sure that I understood these figures correctly. Did the 'median corrcoef median signal' on the right represent the median of the 'corrcoeff of median signal' on the left? Why they were so difficult, even by sign?

L130: "the correlation between AR1 and variance is generally positive for individual synthetic series" – any result supporting this argument?

Figure 3: I must miss something. How to determine the SNR in the real data? And did it happen to exhibit $SNR = 0.1$ and 2 in the empirical data?

L242: what does it mean by 'aggregated'? You had explained that first aggregating time series and then calculating AR1 and variance can remove the influence of changes in satellite instruments to some extent. So did you mean 'first calculating AR1 and variance and then taking the averaging of these metrics'?