

Comment on **essd-2022-54**

Anonymous Referee #2

Referee comment on "*Aridec: an open database of litter mass loss from aridlands worldwide with recommendations on suitable model applications*" by Agustín Sarquis et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-54-RC2>, 2022

The manuscript entitled "aridec: an open database of litter mass loss from aridlands worldwide with recommendations on suitable model applications" describes a database of litter mass loss collected from aridlands across the world. In total, it includes 184 entries from various climate conditions. Then the authors demonstrate how this database can be used to fit decomposition models with different structures. This study collected the litter mass loss data and compiled them into files compatible with R and the SoilR package, which makes it easy to use and expand in the future. Focusing on aridlands and containing time series of litter C loss makes this database unique. However, I have a few questions about its potential usage and the implications derived from this work.

1) The authors argued "a central application of this database is the development of models of litter decomposition for aridlands", and they used a subset of the database (30 entries) to examine the identifiability of various model structures (1-, 2-, 3-pool models, with parallel and series structure for last two). The results show that two-pool parallel and series models with one fixed parameter have the highest identifiable percentage. I'd like to learn more opinions from the authors on whether and how the data samples could affect this result. In these 30 entries, about half of them have more than 5 time points, and only five entries have more than 8 time points. Will including more sites with >8 time points change the result? For the three-pool model, it seems like the series structure has a better identifiability under 2 parameters (Fig 5d and 5e). Is it possibly caused by the entries used (only sites with >8 time points used in three pool series)? If these models are tested by different subsets of the database, how do you justify the identifiability comparison among them?

2) Same question for Figure 6: whether the same or a different number of entries were used for testing these models? If not, the results revealed are not as much as percentage change by using the same denominator (entry numbers), but more like the model collinearity for certain groups of sites. I am not convinced this is an apple-to-apple comparison.

3) in the following applied case study (Figure 7), why do you only select Day2018? Will

the data entry with more time points be able to separate the performance of different model structures? In Fig 7a and 7b, only seven observed time points were included. I am curious why the observed litter mass first declined and then increased to a level close to T0. Is this dynamic caused by more litter input?

4) The applied example is very simple and doesn't show data capability in different models. I doubt the reuse potential unless there are more data with multiple time points (this information seems missing or unclear for the entire database, I only find the entry percentage of time points in the 30 entries). This manuscript could be improved if the authors provide further investigation on 1) how aridland litter mass loss measurement can be improved to better inform modeling development, and/or 2) what modeling efforts can be targeted to change from a data perspective (e.g., missing or underrepresented mechanisms, behaviors of the two-pool model with parallel, series, and feedback structures, etc.).