

Comment on **essd-2021-454**

Anonymous Referee #2

Referee comment on "A compiled soil respiration dataset at different time scales for forest ecosystems across China from 2000 to 2018" by Hongru Sun et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-454-RC2>, 2022

Overall

The dataset is a new compilation of soil respiration rates across China, available from a range of literature. The compilation has an impressive number of values collated. Whilst, by nature, the data is not unique on an individual basis due to being extracted from other sources, the compilation as a whole is a unique resource.

The dataset seems like a potentially useful compilation of values in terms of investigating climate change in China and globally. The values cover a broad range of climatic zones. The work attempts to standardise measurements in terms of different temporal scales, and explains the methods for doing this, which is commendable.

The discussion uses the data to link soil temperature to soil respiration, and thus climate change, and suggests further work could be carried out in relation to soil moisture.

The WEBPLOTDIGITIZER method to extract values from figures seems interesting and an assurance of the quality of these data is given in Section 2.3.

Attempts are evident to show data consistency in collection in Section 2.2 in terms of only choosing R_s values measured from undisturbed ground, and in terms of the instruments used. It could perhaps be explained a little more as to the potential differences that could arise by using different equipment, and how this might affect the dataset.

The article overall is succinct, well-structured and clear.

Data quality

The dataset is easily accessible via the given identifier.

I would normally expect a non-proprietary format for long-term storage/publication of data – e.g. comma separated values (.csv) rather than Microsoft Excel (.xlsx), for purposes of longevity, and to ensure the maximum number of users are able to open the dataset in freely available software.

I would have expected the sample information (Province, Study site, Lat/long etc.) to persist for each data point, rather than there being rows of blank information. The assumption is that samples below the first instance of each Province, Study site etc., are

the same/related, however, if you were to re-sort the spreadsheet, you would lose this associated information from the samples with blanks - each row is not 'stand-alone' as it should be. This also means that it is not clear as to the difference between samples - for example, there are two data points marked with "Aug.,2013" - but what is the difference between the two? There is nothing on the individual rows to explain or describe.

Again related to this, there is no way to automatically calculate the means in order to check their accuracy, because there is no field value by which to group the values to create the mean. The mix of data types in one column also precludes this - e.g. a numeric value column, with "NA" (string/character format) for missing values. I would have expected a numeric code to denote "NA", or a separate column containing the "NA". Error estimates are not given in the dataset, although it is not clear as to whether this would be appropriate, based on the data extracted from the sources. Overall errors are presented in Figures 3 & 4.

Whilst metadata is available in the article under review (e.g. table 1), I would have expected a metadata document (containing field level metadata) to accompany the data download, in addition to the summary given on the Pangaea landing page, which is not very detailed.

Specific Issues

Line 57 - bugedt -> budget