

Earth Syst. Sci. Data Discuss., author comment AC3
<https://doi.org/10.5194/essd-2021-454-AC3>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC3

Hongru Sun et al.

Author 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-AC3>, 2022

Dear Prof. Referee #3:

Thank you very much for your kind consideration and help to our manuscript! According to your suggestions, we revised our manuscript. All the modifications were listed as follows.

Comment: Based on a thorough review of 568 original research articles and other publications, Sun et al. compiled a comprehensive soil respiration dataset that covers a wide range of climates, elevations and forest ecosystems across China. The dataset comprises a total of 10288 monthly and 634 annual soil respiration measurements and for some sites also monthly soil temperature measurements (at 5 and/or 10 cm depth). In addition, specific information (geographic location, forest type, mean annual air temperature and precipitation, etc.) are provided for each site. In view of the vast number of independent soil respiration studies from different regions, consistent datasets that summarise the state of the art and present the available data in a common format are of great benefit for the research community as they facilitate, for example, the analysis of spatial variations and temporal trends in soil carbon emission. The manuscript is generally well-written and the entire dataset is made publicly available through the open-access data repository PANGAEA. However, I have some concerns and suggestions that I would like to see addressed before I can recommend the manuscript for publication.

As I am not an expert in the field of soil respiration, my comments focus mainly on the overall content of the manuscript and the structure of the dataset.

Response: Thank you very much to take the time to review and improve our manuscript and the dataset.

Comment: General comments

Uniqueness of the dataset: A similar dataset as the one presented by Sun et al. has been published by Jian et al. (2020) for different forest ecosystems across China, although it is less comprehensive. I therefore encourage the authors to clearly state the added value of their dataset compared to previous studies.

Response: Uniqueness of the dataset was discussed in the section "**4.3 Improvements of the dataset**" in Lines 260-286, for example, the consistency of the measurement methods, the large increase of monthly and annual Rs samples, soil respiration rates and

concurrently measured soil temperature extracted from the figures with the digital software (WEBPLOTDIGITIZER), the extension of forest types (including Bamboo forests).

Comment: Period of the dataset: I acknowledge the effort of the authors to screen 568 publications and compile all the data, but it would be highly desirable if the dataset could be extended until 2020 so that it would cover a 20-year period (2000-2020), in line with the last two decades of the latest climate reference period (1991-2020). Such a dataset would facilitate the joint analysis of spatio-temporal climate and soil respiration variations (in the context of climate change). Similar to the title for the data repository (Sun et al. 2021), the period of the presented dataset should also be included in the title and abstract of the manuscript. Moreover, it would be helpful to mention the (average) length of the individual time series from the different sites somewhere in the text or to provide a respective figure (e.g. histogram) in the supplements.

Response: Thanks for your suggestions, it took us three years (from 2019 to now) to establish the dataset, it's an interesting and valuable work, we will go on updating the dataset in the future study. Title "Soil respiration database at different time scales in forest ecosystems across China" **was revised to** "A compiled soil respiration dataset at different time scales for forest ecosystems across China from 2000 to 2018".

Frequency distribution histogram of the length of the individual time series from the different sites was supplemented in Fig. S4. "Fig. S4 showed the length of the individual time series from the different sites, the high frequencies were 12 months (38%), 6–7 months (20%) and 13–24 months (15%)." **was added** in Lines 280-282.

Comment: Uncertainties: As mentioned in chapter 2.3, most of the soil temperature and respiration data (82 %) were extracted with the WEBPLOTDIGITIZER. This is an interesting approach that provides a workaround to compile scientific data that are not made publicly available in the original studies. However, I was wondering if the authors of these studies have been contacted to request access to the numeric data or was this not feasible due to the number of studies? Were the data from the 568 studies (many of them non-peer-reviewed) checked manually or automatically to identify potential errors or inconsistencies?

Response: The database included 568 literatures, it was not feasible to contact all of the authors. The data in figures were extracted by Hongru Sun and checked manually by Bingrui Jia. Additionally, the data from the same authors and different sources (e.g. master or Ph. D. dissertation and journal article) has been carefully cross-checked.

Comment: The given R^2 values of 0.99 for the simple linear regressions (original mean soil respiration data vs. digitised soil respiration data) seem promising, but how does it look like for the monthly data? As a measure for uncertainty, the RMSD or MAE should be provided as well. I am also missing a section in the manuscript that discusses (at least qualitatively) the potential uncertainties originating from the different instruments and experimental setups at the different study sites as well as from the varying time periods of the datasets used for characterising differences between the four climate zones (cold-temperature, temperate, subtropical and tropical). Lastly, are there forest types (e.g. mountain forests) that are potentially under-represented in the dataset (due to a lack of respective studies)? Such a potential bias might affect the numbers provided for the temperature sensitivity of soil respiration and for the annual soil carbon emission originating from forest ecosystems in China. This needs to be discussed at least briefly.

Response: Yes, these were monthly data in Fig. S1, because most of the monthly data were shown with figures and the annual data were directly given in the original papers. Root Mean Square Error (**RMSE**) (also called the root mean square deviation, **RMSD**) was supplemented in Fig. S1.

"Rs measurements were mainly from Li-8100 (47%) and Li-6400 (33%), secondary from gas chromatography (18%), and Li-8150 only accounted for 2%." **was added** in Lines 261-262.

"The spanning years were 2003–2014 in cold-temperate zone, 2000–2018 in temperate zone, 2002–2017 in subtropical zone and 2003–2017 in tropical zone." **was added** in Lines 172-174.

"It's worth noting that the Rs studies were fewer in the regions of latitude larger than 48° (~2%) or elevation higher than 3000 m (~4%). The potentially under-represented forest types might affect the evaluation of temperature sensitivity of soil respiration and annual soil carbon emission at the regional and national scale." **was added** in Lines 287-290.

Comment: I would suggest the following modifications for the dataset:

Use a non-proprietary data format (e.g. CSV file) so that the dataset can be easily read by any software.

Response: The format of the dataset was changed to a non-proprietary data format (i.e. CSV file) in the repository in Pangaea.

Comment: Add a metadata file or readme file that contains all necessary information (e.g. those from Table 1 in the manuscript) so that the dataset can theoretically be used independently of the data paper. Nevertheless, add a reference to the data paper in the metadata/readme file and on the landing page of the repository.

Response: A readme file was supplemented in the dataset in Pangaea. There was a reference on the landing page of the repository in Pangaea.

Comment: Create a GeoPackage (.gpkg) or Shapefile (.shp) that contains the metadata (coordinates, elevation, study site name, forest type, etc.) for each study site. Include it in the repository so that it can be easily imported in a GIS by potential users for spatial data visualisation and analysis.

Response: Thank you for your suggestions. We created an ESRI Map Package (.mpk), the metadata was included in the package and was added in the repository in Pangaea. Potential users for spatial data visualisation and analysis can use this ESRI Map Package.

Comment: Add the units (of each column) either in the header or in the metadata/readme file.

Response: The units were added in the readme file in the repository in Pangaea.

Comment: Column "Month": Use the international date format (ISO 8601) or another common date format that can be easier interpreted by a machine (e.g. 2013-07 instead of Jul.,2013). Replace "Month" by "Date".

Response: Column "Month" **was revised to** a common date format (e.g. 2013-07). "Month" **was revised to** "Date".

Comment: No need for column "Period" as the necessary information are already provided in column "Month" ("Date").

Response: Column "Period" means the period of annual soil carbon efflux and is necessary, because some studies only supplied annual data, but not monthly data or only part of monthly data.

Comment: Column "Time": which time is meant here? No time zone provided. It is unclear to which data the time refers. Split into two columns as well: e.g. "Start" and "End".

Response: Soil respiration is usually measured a few days per month, here, "Time" means the observation time per day. "Observation time of Rs" **was revised to** "Observation time of Rs per day (Beijing time)" in Table 1. There were 3 observation times (daily, monthly, yearly) in the dataset, splitting into two columns would be more confused.

Comment: Remove the tilde in "Age" as this special character is difficult to handle during automatic processing. Alternatively include a column before or after and use a flag (1, 0) to indicate whether the "Age" is measured/precise (e.g. 1) or estimated/approximate (e.g. 0).

Response: Age of a natural forest is generally estimated from historical records or dominant tree rings, and the age of a planted forest is defined since planting. Thus, the tilde "~" in column "Age" was deleted, and "estimated from historical records or dominant tree rings in natural forest, defined since planting in planted forest" **was added** in Table 1.

Comment: Column "Forest type": would it be possible to use an integer or acronym for each forest type in the database and provide the full name in the metadata/readme file?

Response: To give the detailed forest community of the study site, forest type was characterized by the dominant tree species, or the ecological similarities (e.g. life form and biotope). The number of forest type was 180, it was not suitable to substitute with an integer or acronym.

Comment: Columns "Rs", "T5", "T10": Better remove "NA" (leave cells empty) and create another column before or after indicating with a flag whether data are available (e.g. 1) or lacking (e.g. 0). The same for the other columns where NA values exist.

Response: The missing values (i.e. NA) in the numeric value columns of "Rs", "T₅" and "T₁₀" **were deleted** in the dataset. A total of 17 columns included "NA" in the dataset. If we created the additional 17 columns, the dataset would become complicated.

Comment: Column "Annual Rs": Does this column indeed provides annual averages or rather the mean over the study period? I think it can be deleted as the mean can be easily calculated from the monthly data provided.

Response: Column "Annual Rs" is the annual soil carbon efflux ($\text{g C m}^{-2} \text{ year}^{-1}$), not the mean over the study period.

Comment: Column "Altitude": Replace "Altitude" by "Elevation".

Response: Column "Altitude" **was replaced by** "Elevation".

Comment: Although the redundancy may increase the file size of the dataset considerably, I would recommend to copy the metadata (geographic information etc.) into each line (not only in the first row from each site). Otherwise, complications may arise when the dataset is reformatted or analysed. Alternatively, the table could be split into two related datasets. One would include the soil respiration and temperature data and the other one the metadata for each site. An additional ID could be provided for each study site to link the two datasets...

Response: With your suggestions, the table was split into two related datasets, one includes the soil respiration and temperature data and the other one the metadata for each study.

Comment: Specific comments

Title: Mention the timeframe of the dataset ("2000-2018" or "2000-2020") and replace "database" by "dataset". A database describes a collection of multiple datasets that are generally stored and accessed electronically from a computer system... Title suggestion: "A compiled monthly soil respiration dataset for [various] forest ecosystems across China from 2000 to 2018"

Response: Title "Soil respiration database at different time scales in forest ecosystems across China" **was revised to** "A compiled soil respiration dataset at different time scales for forest ecosystems across China from 2000 to 2018".

Comment: Line 64-73: Maybe just quote the database here and add the URL (with the access date) in the reference list.

Response: The Five databases were searched the related soil respiration studies to compile our dataset, it was more suitable to directly introduce them in the Section "**2.1 Data sources**".

Comment: Line 75-83: Indicate the period that has been considered. From 2000 until 2018 I think.

Response: "The observation years were from 2000 until 2018." **was added** in Lines 94-95.

Comment: Line 85: Do the 568 publications represent 568 study sites or are some data from the same site? Please state the number of considered sites and their geographic and elevational distribution somewhere in the text.

Response: Some data from multi-sources in the same site, forest type and author were merged in the dataset. "The dataset covers 28 provinces in China (18.61–52.86° N, 84.91–129.08° E)" **was revised to** "There were 155 study sites from 28 provinces in China (18.61–52.86° N, 84.91–129.08° E, 7–4200 m)" in Lines 91-92.

Comment: Line 86-91: Why have the other provided variables (e.g. mean annual temperature and precipitation as well as elevation) not been included in the analysis? I am aware that an in-depth analysis is beyond the scope of this data paper, but some additional plots (soil respiration vs. elevation, or soil respiration along selected temperature or precipitation transects) would emphasise and showcase potential of this dataset.

Response: Thanks for your suggestions. The plots of annual soil carbon effluxes with mean annual temperature and precipitation was added in Figure S3 in supplementary material. "The annual soil carbon effluxes increased with the increasing of mean annual temperature and precipitation at the national scale (Fig. S3). " **was added** in Lines 176-178.

Comment: Line 100: In addition to R^2 , provide the RMSD or the MAE as a measure for uncertainty.

Response: "The Root Mean Square Errors (RMSE) of R_s , T_5 and T_{10} were $0.09 \mu\text{mol m}^{-2} \text{s}^{-1}$, $0.35 \text{ }^\circ\text{C}$ and $0.44 \text{ }^\circ\text{C}$, respectively" **was added** in Lines 104-105.

Comment: Line 105-107: Please clarify whether this sentence describes the procedure in the original studies or how you have modified and analysed the data.

Response: This sentence describes the procedure in the original studies, we didn't modify and analyze the data. To avoid to misunderstand, " The typical days were usually selected to calculate mean monthly Rs" **was deleted**.

Comment: Line 115: I am missing a brief section in the methods about the (statistical) analysis that had been performed to present the data.

Response: The section "2.5 Statistical analysis" **was added** in Lines 122-132.

Comment: Line 119: Provide in addition to the total number of paired measurements (6341 and 2878) also the percentage (in parenthesis) with respect to the total number of considered data.

Response: "There were 6341 and 2878 samples of paired Rs & T₅ and Rs & T₁₀" **was revised to** "The samples of the paired Rs & T₅ and Rs & T₁₀ were 6341 (69%) and 2878 (31%)" in Lines 135-136.

Comment: Line 123-124: Note that R² is invalid/inappropriate for non-linear regressions! R² cannot differentiate between "good" and "bad" non-linear models. The standard error of the regression could be used instead.

Response: "the coefficients of determination for tropical ecosystems (R²=0.225–0.291) were smaller than those in other three zones (R²=0.516–0.934)" **was revised to** "RMSEs in cold-temperate and temperate zones (1.52–1.67 μmol m⁻² s⁻¹) were larger than those in subtropical and tropical zones (1.04–1.32 μmol m⁻² s⁻¹), except the smallest RMSE from T₁₀ in cold-temperate zone (0.42 μmol m⁻² s⁻¹)." in Lines 140-143.

Comment: Line 125-132: Uncertainties need to be provided for these values!

Response: These Q₁₀ values were calculated with the exponential equations between all soil respiration and soil temperature data in each climate zone, but not the means. Additionally, RMSEs of the exponential equations between Rs and soil temperature were given in Figure 2 and Figure S2.

Comment: Line 134-150: One big problem I see here is that time series spanning different years were used to determine seasonal and geographic differences in soil respiration. I am aware that this is unavoidable when using a compiled dataset, but the uncertainties originating from this issue should be at least discussed qualitatively. Which period do the data cover that were used to compute these values (2000-2018?). Add this information.

Response: "which derived from the similar years in cold-temperate (2003–2016), temperate (2002–2018), subtropical (2000–2017) and tropical zones (2003–2015)." **was added** in Lines 151-153.

Comment: Line 149: What do you mean with "winter" in the (sub)tropics? This term does not fit in this context.

Response: "winter" **was revised to** "November–April" in Lines 168-169.

Comment: Line 154-176: Are the data precise enough to state two decimal places? Confidence in these numbers would be increased if uncertainties were provided.

Response: Yes, the annual soil carbon effluxes were precise enough to state two decimal places. With your suggestions, the standard errors were added in Lines 176-202.

Comment: Line 154-155: Are these mean annual values averaged across all study sites? What's the considered time period?

Response: "Mean annual soil carbon emission was $851.88 \text{ g C m}^{-2} \text{ yr}^{-1}$ in China's forest ecosystems, ranging from $260.10 \text{ g C m}^{-2} \text{ yr}^{-1}$ to $2058.00 \text{ g C m}^{-2} \text{ yr}^{-1}$ " **was revised to** "The annual soil carbon effluxes ranged from $260.10 \text{ g C m}^{-2} \text{ yr}^{-1}$ to $2058.00 \text{ g C m}^{-2} \text{ yr}^{-1}$ in China's forest ecosystems, and the mean was $851.88 \pm 12.75 \text{ g C m}^{-2} \text{ yr}^{-1}$." in Lines 174-176. The observation period was not considered.

Comment: Line 160: These acronyms were not defined before. Please specify.

Response: Full names of these acronyms **were added** in Lines 183-185.

Comment: Line 165-176: Too many numbers in these paragraphs. I would recommend to provide a comparative figure instead.

Response: The comparative figure had been provided in Figure 4, and the comparisons of annual soil carbon emissions among different forest types and climate zones were summarized from Figure 4 in these paragraphs.

Comment: Line 177: A general discussion on uncertainties is lacking!

Response: "From Fig. 4 we could also found that the standard errors in tropical and temperate zones ($\sim 16 \text{ g C m}^{-2} \text{ yr}^{-1}$) were smaller those in cold-temperate and tropical zones ($\sim 65 \text{ g C m}^{-2} \text{ yr}^{-1}$)." **was added** in Lines 254-256.

Comment: Line 179-196: A brief comparison with values from other regions outside China could be added.

Response: Our results were compared with global forest ecosystems in Line 210.

Comment: Line 193: See previous comments regarding the use of R^2 and non-linear models.

Response: "The correlations between R_s and soil temperature were lowest in tropical zone ($R^2=0.225-0.291$, Fig. 2d)" **was revised to** "Soil temperature at the depth of 5 cm and 10 cm could only explain 29% and 23% of the R_s variations and RMSEs were $1.09 \mu\text{mol m}^{-2} \text{ s}^{-1}$ and $1.13 \mu\text{mol m}^{-2} \text{ s}^{-1}$ in tropical zone, respectively (Fig. 2d)" in Lines 218-220.

Comment: Line 207-2018: It is difficult to compare any total numbers of R_s (from different sites or studies) if the respective measurement periods, for which these numbers have been computed, are not stated.

Response: The annual soil respiration data in these large-scale syntheses were also from the publications. Due to the limit of annual soil respiration data, it is hard to compare the annual means in the same measurement periods. In this paragraph, we compared the mean annual soil carbon efflux with the previous studies, not the total soil carbon efflux.

Comment: Line 214: I would suggest to write the full names and regret from using acronyms if the terms are only used a few times throughout the manuscript as in this case. This enhances the readability.

Response: The acronyms of forest types were replaced by the full names in Lines 241-246.

Comment: Line 237: This number conflicts with the total number of Rs data (=10288) stated at the beginning of the manuscript, doesn't it?

Response: A mean or a few values per month were shown in the original papers. The samples of 10288 included monthly means and a few values per month. "monthly means or a few values per month" **was added** in Table 1.

Comment: Line 258: This sentence is a bit misleading as no in-situ measurement have been performed. Specify that a comprehensive literature review has been conducted to generate the dataset.

Response: "we collected in situ Rs measurements" **was revised to** "we reviewed the Rs-related literatures and collected in situ Rs measurements" in Lines 297-298.

Comment: Line 251: Any kind of outlook is missing. Does this new compilation for example indicates that there are particular regions or forest ecosystems that are under-represented with respect to soil respiration and temperature studies and deserve more attention?

Response: "It's worth noting that the Rs studies were fewer in the regions of latitude larger than 48° (~2%) or elevation higher than 3000 m (~4%). The potentially under-represented forest types might affect the evaluation of temperature sensitivity of soil respiration and annual soil carbon emission at the regional and national scale." **was added** in Lines 287-290.

Comment: Figure 1: I would suggest to include more information in this map. A digital elevation model, hillshade, orthophoto or land surface cover classification could be displayed as a background map. Different colours or symbols could be used for the study sites to indicate for example the length of the time series (e.g. 1 year, 5 years, 10 years, 20 years), or the number of available variables at each site (i.e. Rs, T5, T10). The overview map is too small and has no added value. Better increase or remove it. It would also be helpful to indicate the considered climates and/or forest types.

Response: Forest types were displayed as a background map, and the average length of the individual time series from the different sites were added in Figure 1.

Comment: Figure 2: Note that R^2 is invalid/inappropriate for non-linear regressions (see previous comment). Have monthly or annual data been used for the calculation? From which period do the data originate?

Response: Root Mean Square Error (RMSE) was added in Figure 2. Monthly soil respiration rates and soil temperature from 2001 to 2018 were used for the calculation.

Comment: Figure 4: How and for which period were the mean annual fluxes calculated? I assume all annual data from different years and sites associated with one forest type were spatially and temporally averaged. Is this correct?

Response: Yes, all annual data of the same forest type in each climate zone were averaged.

Comment: Figure S1: Does the mean reflects the entire study period at each site? Do the correlations look similar if monthly data (collected vs. digitised) were compared?

Response: Yes, the mean reflects the entire observation period at each study. The correlations (collected vs. digitized) were excellent and R^2 were all larger than 0.99.

Comment: Figure S2: I have the impression that there are other non-linear functions that describe the relationship between the soil respiration rate and soil temperature better than the applied ones.

Response: Figure S2 showed the relationships of soil respiration rates with soil temperature at 5 cm depth and 10 cm depth across China, Figure 2 showed the respective relationships in four climate zones.

Thanks again for the reviewer and the editor for your kind consideration and help!

Best regards

Sincerely yours,

Hongru Sun, Zhenzhu Xu, Bingrui Jia