

Earth Syst. Sci. Data Discuss., referee comment RC1  
<https://doi.org/10.5194/essd-2022-384-RC1>, 2023  
© Author(s) 2023. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on **essd-2022-384**

Anonymous Referee #1

---

Referee comment on "The Ant-Iso dataset: a compilation of Antarctic surface snow and ice isotopic observations" by Jiajia Wang et al., Earth Syst. Sci. Data Discuss.,  
<https://doi.org/10.5194/essd-2022-384-RC1>, 2023

---

In the manuscript "The Ant-Iso dataset: a compilation of Antarctic surface snow isotopic observations", Wang et al. present a compilation of published and unpublished isotopic data ( $\delta^{18}O$ ,  $\delta^2H$  and D-excess). The manuscript is largely well-written and follows a logical structure, but different points were unclear. For instance, it was explained that the "MD08" isotope data was supplemented with 794 newly collected data (L89-93). However, it was not explained:

- which methods were used to find the different studies
- which keywords were used
- which time scales were considered
- why were certain studies in or while maybe others excluded
- why specific water sources were included or excluded
- sampling method
- which data quality parameters were used to decide to in- or exclude
- why certain meta data was presented while others not
- which data was collected by the authors and which from literature (in database)

In addition, it was not explained (database and text) in which way the different spatial and temporal data, and the model time scale were aggregated.

It seems that in the presented data ([doi.org/10.5281/zenodo.7294183](https://doi.org/10.5281/zenodo.7294183)), figures, and tables, all data was shown at once without distinguishing between different spatial resolutions (space and vertical resolutions), temporal resolutions (years and time single vs time series), and various water sources. It would be good to present all "raw" collected data from literature in the database and explained in the manuscript how the data was averaged (e.g. Pang et al. 2019, snow pit 3 m 29 average values; Münch et al. 2017, snow pit 3.4 m 1329 average values 2014-2015). In addition, discuss the value of such averaged data and whether the collected data can be used to compare field data and model results consisting of different time scales. Providing the full collected data allows the user to use the data for different purposes while detailed information on sampling time and statistical weighing in the different statistical analyses, tests, and figures will help to justify whether patterns or relations presented in figures or table hold or whether these patterns appear by chance.

In addition:

L32 Clarify what the difference between a data point and location.

L35 Specify already here geographic and climate information.

L52 Include reference to literature.

L56 Clarify what "information and high fidelity " means.

L61 and L63 "spatial linear" remove spatial in L61 since the emphasis was on the core to avoid confusion with different cores located at various locations.

L62 and L62 Include references at the end of the sentence.

L65 "uneven" Clarify what uneven means.

L70 "high-resolution" Specify resolution.

L74 Include reference to literature.

L77 "high altitude" not consistent use of altitude vs. elevation.

L78 Include a figure to highlight the distribution.

L79 "researchers", which ones?

L89 Explain how the data was sourced.

L91 "strong support" clarify who gave support, the funding agency or colleagues?

L96 Give a brief introduction of "travers sampling" not everyone might be familiar with this type of sampling approach. Include also other sampling approaches and sampling methods and devices used.

L97 Specify who collected this data, the authors or different authors from the literature.

L106 "we added two routes" Clarify whether own data or literature sourced data was used.

L108 "unreleased" Explain what unreleased mean here since the data seems to be published by Ekaykin et al. (2012). Please clarify.

L128 Explain what reliability and its context means.

L129 Explain what a seasonal bias is.

L132 Since the data was collected in the dry valley with little precipitation, the data could still be a multi-year average. Please clarify.

L137 State clearly why only the last few decades were included and not all.

L146 Please also include additional geographic factors such as slope and aspect ...

L157 Add reference to original authors.

L196 Specify which correlation coefficient was used.

L223 Clarify how the comparison was performed (not described in text L190-onwards). It should be included in the method section.

From L190, the model was run from 1979-2018, but the data had different temporal resolutions. Is it fair to compare these data, and what can we learn from this non-equal comparison?! Please clarify and discuss.

L268 scientists who have provided help and support in data collection. These scientists should be explicitly stated and acknowledged in the manuscript.

Table 1 Explain what are sufficient number of measurements are and why and based on what the threshold of 10 was chosen.

Figure 1 c choose different symbols similar to fig 4 for data from various data sources.

Figure 4 Rearrange panels with column 18O and one column D-excess. Include different symbols and colors indicating different time scales.

Figure 5 is hard to read too much data and color scheme, and symbols.