

Geosci. Model Dev. Discuss., referee comment RC3
<https://doi.org/10.5194/gmd-2021-77-RC3>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.



Comment on gmd-2021-77

Anonymous Referee #3

Referee comment on "Cosmic-Ray neutron Sensor PYthon tool (crspy 1.2.1): an open-source tool for the processing of cosmic-ray neutron and soil moisture data" by Daniel Power et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2021-77-RC3>, 2021

The manuscript presents a new processing tool for stationary cosmic-ray neutron sensors based on a python package on GitHub. The tool is capable of reading in CRNS time series data and soil sampling data in a given data format, of filtering and correcting the data, and of generating an output that includes data products like soil moisture, uncertainty, and penetration depth. The tool is also capable of consulting external data, like ERA5, to support the gap filling and meta data description. The author's vision is that processing steps should be harmonized across all CRNS networks and that the community of users and researchers will use and maintain this code to generate their data products.

I do fully support this vision and I agree that it is about time to provide researchers and users a tool to more efficiently and consistently work with CRNS data. Crspy is one of the first open-source tools that offers a timely and substantial contribution towards this goal and hence it is worth to be published in GMD.

General concerns

- The manuscript is concise and well written, but my impression is that it falls short of more elaborate explanations regarding (1) the technical details how Crspy works, and (2) regarding user guidance.
- From a GMD paper I would expect that every single equation and processing step is explicitly described and mathematically clear. This will allow users to fully understand what the model does without looking at the code. Hence, I strongly suggest that all these parts – e.g. about the air pressure correction, the soil sample calibration, or the temporal aggregation, to name just a few – should be much more elaborated. Essentially, it would require not much more than typesetting the procedures used in the code. But from my understanding this is standard for articles on new tools and models.
- An important detail which directly follows from the previous comment is that it was not clear to me from reading the manuscript how aggregation and/or smoothing of the data

is performed. Do you aggregate neutrons before conversion to soil moisture, or do you aggregate the final soil moisture product? Is the aggregated data indexed at the start, middle, or end of the aggregated period? These details sound picky in the first place, but they are of major importance since they can have substantial effect on the final product (due to the non-linearity of $\theta(N)$) and on the comparability to other processing tools.

- I would suggest that the manuscript elaborates a little bit more on the details of how crspy works internally and how it should be maintained by the community. Not because the needs of expert programmers should be addressed, but rather to facilitate community-driven updates of the code. Since the CRNS research changes their methods often, it would be a key feature of crspy to be adaptable by the community. So please provide a key section on (1) guiding researchers how the code could be changed, e.g., if a new correction function needs to be included, and (2) guiding users what to do if they want to use the new correction (update the script, change meta data, etc.). Add also discussion on how can the community make sure that scientists regularly update their code? How can users of the data verify the the processing scheme of a data is up to date?
- My impression is that the authors undersell Crspy in this short manuscript. It looks like crspy has a lot of useful features and products, which are only marginally mentioned in the text and figures. I would suggest to more prominently illustrate potential data products of crspy, e.g., a soil moisture time series including their error band, the footprint depth, examples of flagged data in certain periods, or diagnostic output. Moreover, it is very promising to see that the metadata can be used to do meta analysis on the data, but you only show examples using land use or meteorological data. From my perspective, the meta data analysis would be even more valuable for the CRNS community when looking at site-specific parameters, soil properties, and their correlation to N_0 , GV , or biomass, for instance. I'd recommend to also provide such an example (similar to what was used in Shuttleworth et al. 2013 to correlate COSMIC parameters with soil bulk density), as this would push the community research a lot forward.

Technical concerns

- Naming conventions (L124: "it is first necessary for a user to correctly format input data following crspy's naming convention"): I think one of the biggest obstacle for users to apply the new tool would be that crspy requires a certain data format (Tables A1 and A2), while most data portals and CRNS data providers have already their data format fixed. I'd suggest to slightly adapt the crspy meta files such that the user may define the column name of their variable individually (e.g., temperature_column name = TEMP. Then, crspy can address the data by the given column name, independent of column position or other restrictions, which would facilitate much smoother integration into existing data workflows.
- The initial setup of the code requires a steep learning curve and a lot of prior knowledge. There might be ways to lower the bar for users, e.g.,
 - by providing a setup wizard script which assists the user in setting up their first connection to the data base and creation of a station and its meta data.
 - by providing a first fully-working example with which the user could start right away after installing the tool.

- by deploying the project with Docker, a platform that unfolds automatically on any system without prior trouble regarding python installations.

Minor concerns

- Eq. 9: Crspy cuts off soil moisture at the maximum porosity of the site. Does that mean that crspy cannot be used for periods of snow or ponded/intercepted water?
- ERA5: while it is a good idea to fill data with ERA5, there could be a spatial mismatch of scales and also a bias in absolute values due to metrological reasons. This could lead to a significant bias of neutrons particularly when gap-filling air pressure. Is there a way to first compare local data with ERA5 data, identify their constant bias, and then use it to gap-fill? At least it would be good for crspy to create diagnostic output showing local air pressure versus ERA5 air pressure (and humidity, temperature, ...) in order to check the consistency from time to time.
- Line 258: When the option to find the nearest NM is used, why do you ignore the GV-correction factor? It could still be that the nearest station is many GVs away. And, what if the nearest station has no data? (as it often happens on nmdb.eu), do you take the next nearest station?
- Filter: You seem to remove counts below 30% of N_0 , which is reasonable as one would not expect a stronger effect on the neutrons than pure water could have. But I do not agree on removing counts above N_0 . Under very dry conditions, counts could exceed N_0 by a few percent. (check $N(\theta=0, N_0) > N_0$)
- Filter: very often the incoming correction does not work very well in periods of ground-level enhancements from cosmic-rays or during coronal mass ejections. One could offer another option to flag also data where the change of neutron monitor data is suspicious, i.e., drops by a few percent from one day to another.

Specific comments

- Eq. 1: The authors may want to acknowledge the new formulation from Köhli et al (2021), which is equivalent to the function from Desilets et al. (2010) using $N_{\max} = 1.075 N_0$. It would not change anything in the results. But the Desilets equation has 4 parameters, which is one parameter more than necessary and thus the function is overstated and would lead to non-unique calibration results. The equivalent reformulation is only 3 parameters and would lead to unique optimization results if needed. Note that the other UTS equation proposed by Köhli et al. is too new to be addressed by this version of Crspy, but it is a good example of quickly changing CRNS methods, and demonstrates how important it is that Crspy should be flexible and adaptable by the community.
- Line 73: citations of networks are:
 - Zacharias, S. et al. A Network of Terrestrial Environmental Observatories in Germany. *Vadose Zone J* **10**, 955–973 (2011)
 - Bogena, H. R. TERENO: German network of terrestrial environmental observatories. *J Large-scale Res Facil Jlsrf* **2**, 52 (2015).
 - Cooper, et al.: COSMOS-UK: national soil moisture and hydrometeorology data for environmental science research, *Earth Syst. Sci. Data*, 13, 1737–1757, (2021)
- Line 102: Please reformulate "currently this revised approach is not applied across the networks", I think many recent studies adopted the new approach, and the COSMOS-UK network seems to apply that the new weighting, too (please double check)
- Parts of Fig 1 are hard to read, can you increase font size (e.g. reduce time stamps and

- number of rows in the boxes) and increase resolution (e.g. vector format, pdf/eps)?
- Line 140–145: note that the equation has been shown to work not very well for very dry conditions (Köhli et al. 2021) and that the parameters $a_0..a_2$ are not always constant for all sites, since literature exists where these parameters have been adapted. Since crspy aims at offering a general solution, would there be a possibility to change this equation or to automatically fit these parameters?
- Eq 2: can you please provide a reference for this equation and the 0.556 factor?
- Line 170: if LW is not provided, why don't you estimate it from clay content using the soilgrids data base?
- Line 223: The reference from Desilets 2021 is only a technical document and it is not available under the given doi. Please check whether a better reference could be given or provide details of the exact calculation. E.g., if it requires cut-off rigidity where do you get that number from?
- Line 245: Please clarify again the difference between the two correction factors f_i and f_i' for incoming intensity.
- Filter: When do you apply the filters, before or after correction of N_{raw} ?
- Filter: very often sensors have maintenance periods where the data is not to be trusted. Does crspy support manual definitions of to-be-excluded periods?
- Line 353: this sentence has a circular problem: "Each site is also be given a country code and a site number in the metadata, which is used by crspy to find any required values stored in the metadata"
- Figure 2: hourly time steps are hard to visualize for long time series given the high fluctuations, and monthly aggregation looks very abstract. I'd recommend to use daily aggregation for all panels. This would allow for better illustration of the differences between the models in panels a–d, and it would no longer require panels e–f.

Appendix

- important note: require date time to be in UTC
- explain what unmoderated and moderated means
- What if sensors have more than 2 or less than 2 pressure sensors?