Response to reviewer #3

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.

We thank the reviewer for taking the time to review our manuscript and for their positive feedback. Their comments and suggestions provided will help us to improve our manuscript in the next iteration. We are also pleased that they also share in our vision to provide open tools for the CRNS community.

General concerns

1. The manuscript is consice 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.
2. From a GMD paper I would expect that every single equation and processing step is explicitely 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.

   In producing this manuscript we wanted to ensure that it is informative and readable and so we have described the key equations that will impact the outputs. Although we agree it is important for users to understand all aspects of the way a model works, if we were to typeset every single equation, we feel it would make the paper unreasonably large. In addition, many implementations of routines are from established literature, and we feel to be sufficient to point the general GMD readership to the appropriate papers without reproducing the same set of equations here, hence focusing on the

integrating aspect of crspy as a data processing tool. We agree however that we should ensure it is clear on any decision points made, such as aggregation methods, to give users a clear understanding of how the data is processed. We will address this in the updated manuscript.

3. 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.

   We thank the reviewer for raising this comment which we believe I might have been due to lack of clarity in our manuscript. Currently we aggregate the soil moisture estimates after processing (i.e. each hourly reading is processed to give a soil moisture estimate, any aggregation is then undertaken on the soil moisture values with the indexing occurring on the end of the aggregated period). For the smoothing we have employed a moving 12-hour average (with a constraint that there is a minimum of 6 hours of data in the window). This 12-hour window is well established in the cosmic-ray neutron sensing community (e.g., COSMOS) and also confirmed by our preliminary analysis at our UK sites (possibly one of the most challenging ones due to soil wetness, proximity to sea level, and humid atmosphere) showing 12-hour averaging to satisfactory (data not shown). We will be clearer about these points in the updated manuscript.

4. 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 dicussion 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?

   This is a great idea and something we will address in the updated manuscript. We have already included on the github page an example

workflow that is intended to help users understand how to run crspy (from installation through to processing). We have also spent time ensuring that the code is thoroughly commented. Including a key section as described above would help to bridge the gap for those who wish to use and alter crspy in the future.

5. 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 prominantly 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 paramaters, soil properties, and their correlation to N0, 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.

   We thank the reviewer for this very positive comment. When writing the manuscript, we have tried to balance demonstrating all the features whilst ensuring it remains concise and readable. We will improve on this in the next iteration through additional figures that highlight some of the features mentioned above.

   Currently, we have initially addressed the main comments made by the reviewer. We will respond to all technical and more specific comments raised by the reviewer should a revised version of the manuscript be invited by the GMD Editor.

   We once again thank the reviewer for the very positive feedback and their comments and questions.