

We thank the Reviewer for the comments. Our detailed replies are in blue below.

RC3: 'Comment on essd-2025-554', Referee #3: Brian Henn, brianhenn@allenai.org

General comments:

The authors describe a dataset of corrected precipitation, which is derived from the Stage IV 4km/hourly precipitation dataset that has been modified to better fit local gauge observations (where available) and streamflow observations using an inverse modeling approach. The dataset spans basins in the Appalachian mountains, while most of the calibration and methodological development has been done in a set of testbed basins in the Southern Appalachians. The authors describe the methods of precipitation dataset correction, and evaluate the resulting precipitation datasets against several metrics.

The resulting dataset may be useful for understanding topographically derive precipitation variability and bias in conventional datasets. Streamflow and local gauge observations can drive improvements in understanding of precipitation pattern that may not be caught the standard gauge/radar network. As a result, this type of study is useful to help improve the observational record, and to understand

However, the manuscript accompanying the dataset struggles, at times, to reliably describe the dataset itself, and to reproducibly describe how the dataset was created. There are many instances of important aspects of the dataset not being consistently described in the high-level description (abstract, introduction) vs in the more detailed methodological sections (which run over 30 pages): eg, its spatial and temporal extent, its degree of validation against independent data, the appropriateness of generalization the method from where it was tuned to other basins, etc.. At many times the methodological details are missing or hard to follow, and rely on the reader having read at least 3-4 other papers that perhaps more completely describe the methods (I have not read those papers so I cannot say the extent to which this is appropriate). Additionally, the code generating these datasets does not appear to have been publicly shared alongside the data, further making reproducibility difficult.

We thank Dr. Brian Henn for the general comments. In the revised manuscript, we added more high-level descriptions as well as more methodological details in response to this

comment. The codes to read the data are provided in the same Zenodo repository as version 4. The hydrological model itself is in the process of licensing and will be made public upon successful licensing.

Specific comments:

L162: “ StageIVDBKC” This is a very long acronym – please spell out fully on first reference.

The authors added details explaining the acronym in Lines 171-173. D is for Downscaled; B is for Bias Correction; K is for Kriging; and C is for Climatological Correction.

L188/Eq. 3: I don’t think this is the equation for power spectral density – it seems to be the equation for computing the Fourier transform of the rainfall field. The PSD is given by something like Eq. 4.

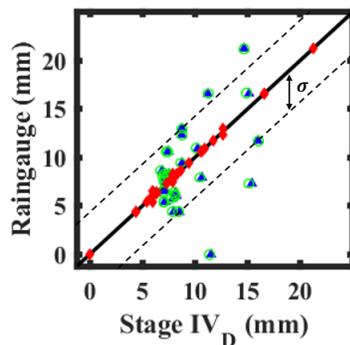
Thank you. The sentence is revised and corrected. Line 197 in the revised manuscript.

L210: What is “ND”? Define.

ND refers to ‘Number of Downscaled samples’. This is now explicitly stated in Line 223.

L226: “A linear regression was applied to StageIVD pixels within one standard deviation of the regression line at an hourly timescale by assuming homogeneity of variances or homoscedasticity”. I can’t follow the meaning of this sentence. What does “within one standard deviation of the regression line” mean? What happens to pixels that are outside this line? A figure illustrating the stages of the downscaling/bias correction using actual data would be immensely helpful here.

‘Within one standard deviation’ means that radar rainfall is not significantly underestimated or overestimated. Radar pixel rainfall that is outside one standard deviation with respect to rain gauge measurements is considered an outlier and thus is not suitable for the linear adjustment using Eq. 8. An example showing one standard deviation envelop (denoted as dashed dark lines) in a scatter plot is below:



L232: “miss detection” should be “mis-detection”

It should be missed detection. It is now revised.

L236: I don’t understand from this description how the method of Light Rainfall Correction is different from Medium Rainfall Correction. Also, these methods appear to be applied at the event scale, but the topic sentence of this paragraph says “The second phase of bias correction is done at decadal scale”. Again, this section is hard to follow and figures showing specific example via maps of precipitation at each stage of the bias correction would be immensely helpful.

These methods (LRC, and MRC) are done for the entire duration of the dataset (2008-2017). They refer to the climatology corrections. We didn’t include these intermediate results in this section because they were published as a report in the Duke Repository in 2019 (Report EPL-2019-IPHEX-H4SE-4). An improved Section 2.3.3 is provided including relevant references.

L258/Eq. 9/Eq. 10: Please provide definitions for gamma, Z, G, etc. in these equations. Also please define what is meant by the Kriging target “interpolates differences between radar data and rain gauge observations”: How is this difference defined? Time-mean? Additive? Multiplicative? It is not possible to reproduce the method from the text as currently written.

Z and G are interpolated differences at any pixel, and actual differences at gauge locations, respectively. The difference was defined earlier in Lines 261-263, which is calculated as the difference between the rain gauge measurement and the radar measurement at hourly resolution at the pixels where gauges are located. It is now revised.

L292: Define HSS

It was defined in Line 290 along with TS and other metrics.

L310: delete colon after “and” and remove hyphen between IRC and improved

It is done.

L439/Figure 4: If there are 28 basins, why are they labeled as basins 1-30 on the map?
Correct to make consistent

The labeling process included all studied basins, but 2 basins were dropped out for subsequent analysis due to the size of the basin being too large or too small. This was detailed in Liao and Barros 2025a. An explanation is now added in response to this comment in Lines 449-452.

L433: Up until now, no distinction has been made in the text between the 28 basins (throughout the entire eastern US) that are the focus of this study and the Southern Appalachian basins on which the IRC method was trained. I thought that they were the same, i.e., that this study was just about the southern Appalachians, but that is not correct. The authors should make this distinction clearer in the abstract and introduction.

Thank you. We regret the confusion. We added a statement to address this issue in the introduction in Lines 90-92. Although the IRC was first demonstrated for basins in the Southern Appalachians in Liao and Barros (2022), it can be applied generally.

L455: “Flood-producing events have been selected for the 28 headwater basins for recent years from January 2021 to April 2024”. But the abstract says “we present a high-resolution (i.e., 250m, 5-minute-hourly) QPE dataset for 215 extreme rainfall events occurred in 26 gauged mountainous basins in the Appalachian Mountains from 2008 to 2024”. Are these different? Why? I am having trouble following the organization of the paper and the study data.

The presented events (215 extreme events for 26 gauged basins in Zenodo) are part of the 28 studied basins. However, due to the Karst terrain (discussed in Lines 578-592), subterranean runoff processes are important in Basins 13 and 14. Because these processes are not represented in the hydrology model, the IRC results for 2 basins are not as skillful as elsewhere, and thus the storm events for these two basins are not included in the final data set available in Zenodo.

L480: “In the entire study domain, rain gauges are only installed in the Southern Appalachians, specifically in the vicinity of the Cataloochee Creek Basin (Basin 05). However, the rest of the regions are not equipped by raingauge networks, and therefore, no rain gauge bias correction is done for those basins, and the downscaled original dataset StageIV (i.e., STIVD) is used as input for the IRC method and hydrological simulations in this study.” This is critically important for readers to understand about this study/dataset. It needs to be stated much earlier than 20 pages into the paper, i.e., in the introduction/abstract. How appropriate is it to ignore model uncertainty when transferring the hydrologic model from basins in which it is calibrated (Southern Appalachians) to completely different regions where it may not be generalize well? Some discussion of this is needed.

A statement addressing the availability of rain gauges and the generalizability of the IRC framework is added in Lines 92-96. Liao and Barros (2023) showed that the IRC’s

performance when rain gauge bias correction was not done as a prior step compares well with when rain gauge correction was done ahead of IRC, which suggests the wide applicability of IRC across regions either with or without rain gauges. The summarized result was highlighted in Figure 12 in Liao and Barros (2023).

L555: “225 events”. The abstract says 215. Which is correct?

Later in this paragraph, Line 589 stated that 10 events from Karst terrain were removed from the published data due to the complexity of Karst terrain and the use of a hydrological model without an embedded Karst module for the IRC. Nevertheless, we decided to discuss and show events affected by Karst terrain to illustrate the importance of getting the right results for the right reason. Because the hydrologic model is not calibrated for the IRC, the poor performance in Karst terrain is expected since the model lacks the physics to describe rainfall-runoff processes.

Fig. 10: It is useful to see an actual example of the rainfall fields during the different stages of processing as are shown here. I would find it useful to show more of these examples when explaining the downscaling/bias correcting/IRC/ICC stages in Section 2.

Thank you. The examples regarding different stages of downscaling and bias correction were published before as a report (Report EPL-2019-IPHEX-H4SE-4). The focus of the paper is on the IRC and ICC and examples shown in Figure 10. To address the Reviewer’s comment, we added the following figure from the report into the Appendix A as Figure A1 to show the impact of different procedures in the downscaling and bias corrections on precipitation fields.

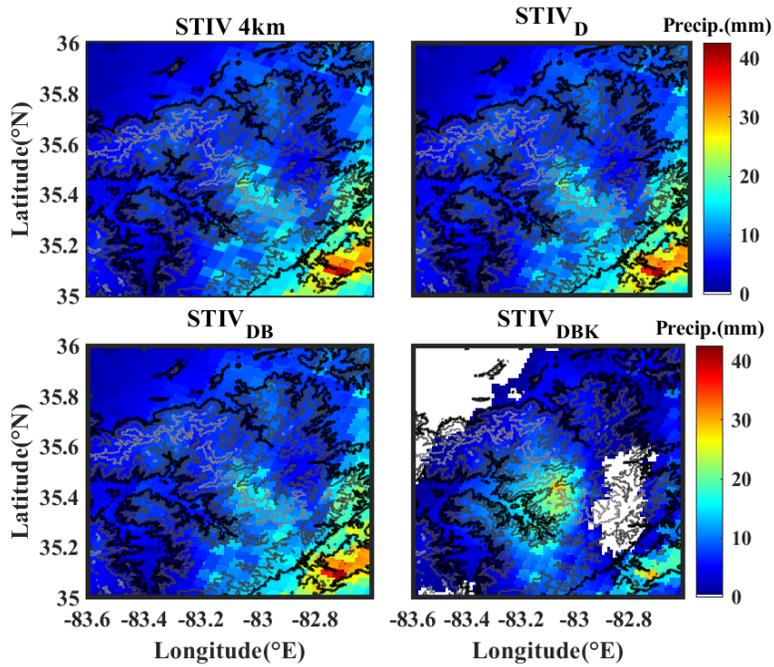


Figure A1 - Spatial rainfall fields on 2014-05-15, 06-07 UTC. Rain rates between 0 to 1mm/h are mapped in white.

Thank you.