Author's comments in reply to referee 2

We thank the referee for the valuable comments and suggestions, which gave us useful insights to improve the quality of the manuscript.
Below we provide a detailed response to all comments. Referee comments are in bold, and author's comments are in *blue and italics*.

Jose P. Teran & Jhon Vila,
On behalf of the authors.

General comments:
This study compared different gridded precipitation data and downscaled precipitation results with ground-based station data. It can be seen that the study area is characterized by data scarcity and complex mountainous terrain, and this research has positive implications for hydrological studies in data-deficient regions. Precipitation observation and simulation in mountainous areas have always been challenging, and cloud and fog deposition are also major sources of precipitation in such regions, yet this component is often difficult to observe. I have several questions regarding this paper and would like the authors to address them.

Specific comments:
1. The main content of this paper is to compare different downscaling method optimized remote sensing precipitation products with station data based on various indicators. The results show that the best product is Rain4PE. Why did the authors not first evaluate the different products before performing downscaling? I think that the overall logic of the paper is not sufficiently clear.

We thank the referee for this question. RAIN4PE covers Peru and Ecuador, while PISCO covers Peru, both regional, gauge-corrected products, whereas GPM (IMERG-F) is a global product that, while gauge-corrected, is not adjusted to the same regional density and detail as RAIN4PE or PISCO. Our aim was to test whether NDVI-based downscaling can improve a global product like GPM, so that the approach is applicable beyond this region, and the NDVI-based downscaling alone adds the most value. RAIN4PE and PISCO were used as benchmarks to evaluate how close the NDVI-based downscaled GPM gets to a regionally-corrected product, not as candidates for downscaling themselves. We agree this might not be sufficiently clear, which we will make sure of in the adjusted manuscript.

2. I am somewhat confused about the structure of this paper. Do the authors intend only to compare different products in mountainous areas, or do authors aim to highlight a precipitation product suitable for mountain regions? Based on the Introduction, I think the authors are more inclined toward obtaining a high-spatial resolution precipitation dataset. If the goal is to indicate which high-resolution product is more suitable for mountainous areas, should not all products be downscaled before comparison?

We thank the referee for the comment. We intend to assess how NDVI-based downscaling methodologies perform in an Andean mountainous region, with the goal of establishing a downscaling method applicable to this and similar mountain regions, rather than comparing multiple existing precipitation products to identify which one is best suited for our specific region. As noted in our response to the previous comment, we applied this downscaling to GPM specifically because, as a global product without the level of regional correction present in RAIN4PE and PISCO, it best represents the case where downscaling would add the most value. RAIN4PE and PISCO were used only as reference datasets to evaluate the accuracy of the downscaled GPM product. We will revise the Introduction to make this objective clearer.

3. In Section 3.3, the authors mention that a threshold of 2.5 mm d^{-1} was set to distinguish precipitation events in order to reduce sensor noise. However, in Section 3.4, the threshold for Dry Days given in the ETCCDI indicator table differs from that used in the paper. When calculating the Dry Days indicator, did the authors use the 2.5 mm d^{-1} threshold or follow the standard threshold specified for that indicator?

That is a great point and we thank you for raising it. We defined the 2.5 mm d^{-1} threshold as a day with rainfall specifically for the rainfall detection/forecast indices, because of the definition by WMO mentioned on the manuscript, but also because of the noise close to zero on the datasets that could lead to incorrectly defined POD or FAR values, which was not so influential above this value. On the other hand, the 1mm threshold is indeed the standard definition from the ETCCDI that we used for the calculation of Dry Days. We do think that these can be inconsistent definitions for a similar concept. Therefore, in the revised manuscript, we will adjust the calculation of indicators with a common threshold that allows to filter out the noise but is still a meaningful value to determine occurrence of rainfall and detection of dry days.

4. Currently, the accuracy of precipitation at high altitudes remains problematic (Miao et al., 2024). Therefore, improving precipitation accuracy in high-altitude regions is an ongoing challenge, and it is unclear whether this issue also exists in the authors' study area. The authors have mentioned some correlations between the products and observational data. I suggest that including relevant discussion on this aspect would make the paper more rigorous.

We thank you for the suggestion. That is indeed also a challenge in the Andes. Components such as wind, and the technology of measuring devices have been shown to have an impact on accuracy in similar regions (Padrón et al., 2020), as well as human error sources and lack of data quality control procedures (Hunziker et al., 2017). Despite our study not focusing numerically specifically on this issue, we will expand the discussion to also consider these aspects, which have implications on the significance of our results.

Minor comments :

1. In line 40, there appears to be a missing space in 'agrid', please correct it to 'a grid'.
2. In line 154, there appears to be a missing space in 'tothe', please correct it to 'to the'.
3. The font size used in the figures appears somewhat smaller than that of the main text; it would be advisable to increase it appropriately for better readability.

Thank you for these comments. We will correct these sections on the revised manuscript accordingly, and will adjust the font sizes on figures to increase readability.

References

Hunziker, S., Gubler, S., Calle, J., Moreno, I., Andrade, M., Velarde, F., Ticona, L., Carrasco, G., Castellón, Y., Oria, C., Croci-Maspoli, M., Konzelmann, T., Rohrer, M., & Brönnimann, S. (2017). Identifying, attributing, and overcoming common data quality issues of manned station observations. International Journal of Climatology, 37(11), 4131–4145. https://doi.org/10.1002/joc.5037

Padrón, R., Feyen, J., Córdova, M., Crespo, P., & Célleri, R. (2020). Rain gauge inter-comparison quantifies deficiencies in precipitation monitoring. La Granja, 31(1), 7–20. https://doi.org/10.17163/lgr.n31.2020.01