Comments to the Author

Review of essd-2023-63: 'MOPREDAScentury: a long-term monthly precipitation grid for the Spanish mainland' by Santiago Beguería, Dhais Peña-Angulo, Víctor Trullenque-Blanco, Carlos González-Hidalgo, submitted to be published in ESSD: Earth System Science Data.

Recommendation: Reject or Major Revision

General Comments:

This work describes the development of a new monthly precipitation dataset for the mainland Spain which is a natural continuation of the previous studies of the authors. The period covered by this dataset has not precedents in this region being of great interest to analize trends and the effect of the different uncertainty sources affecting the interpolation process on time (station's density, spatial coverage, temporal coverage, etc..). In this sense, the study is very interesting and could be a good contribution to the climatic community of the target region. However, in my opinion the work should clarify some points and/or properly discuss and support some affirmations the authors have made before to be considered for publication in ESSD.

Based on these concerns I would recommend a minor revision.

Minor Comments:

Abstract: Which is the dataset referred to by the authors? It seems to be the observational network of NCDB-AEMET but is confuse in the manuscript. Based on your summary, the interpolation process has not two phases alone, those are only referring to the dual nature of precipitation, but more, at least to obtain the monthly precipitation from the monthly anomaly and the monthly climatology.

Introduction: "These efforts are particularly needed in regions where water is scarce and limiting resource combined with a high demand…." I would disagree with this affirmation. In my opinion, in this sense is more relevant the spatial variability of the precipitation and the heterogeneous regimes in the region considered than the water availability as to properly characterize the precipitation in an heterogeneous area you need as much observations as possible. In particular, in addition to the water availability problem pointed out by the authors, the region considered has a great spatial variability driven by the orography. **Line 42:** What are the authors referring to with "models"? Numerical models or statistical models? In the first case, the use of gridded data is mostly due to the areal-average representativity, as the models, in contrast with the point representativity of local observations. **Line 50**: Several more recent datasets have been ignored or not included in the text. In particular, at least SAFRAN-Spain (Quintana-Seguí et al. 2016, 2017), Iberia01 (Herrera et al. 2019) and AEMET-Spain (Peral et al 2017), that have more recent periods and higher resolution than the ones cited by the authors. **Lines 57-60:** The reference for the observational datasets used have changed with respect to the abstract. Is there any reason?

Data Rescue (yearbooks): Line 80: These two sentences seem to be contradictory. I mean, there is a manual quality control but the the observations are automatically flagged as suspicious. There is a bold text and a reference to Table 1 but seems not being the caption of this table but normal text.

Evaluation: Although coherent, the notation is a litle confuse in my opinion. How do the authors define the indicator function? Usually, the 1 corresponds to the occurrence of precipitation and, then, the true positives are defined as obs > 0 and pred > 0. However, the authors use the opposite notation so I suppose that the indicator function identifies the days with 0 precipitation. This is in agreement with your equations, simply I comonly use another definition. Which is the range of the different parameters defined? I mean, PPV is lower than 1 so high values correspond to values close to 1. Similarly, the others parameters have their own range and interpretation.

Results: Section 3.1,: The number of stations is referred to monthly data? The nature of the data is relevant to properly interpret the added value of the yuearbooks with respect to the use only of the BNDC. It is surprising to me that, including in the modern years, the yearbooks could have this clear added value with respect the BNDC. The result about the mean distance to the closest station is reflecting that there is not any station isolated more than the dataset is spatiallu homogeneous. For example, suppose you have a dataset with 6 stations, three on the north and thre on the south of the Iberian Peninsula. If you remove one of each region, the mean distance doesn't change significantly, which is the result you obtain. I mean, is the reduction of the stations the one that is spatially homogeneous, avoiding any isolation of the stations. The effect of the stations reduction is more visible on the increment of the variability of that distance. The correlation between precipitation and orography is high when monthly of climatological values are considered but it is not so evident at shorter time scales. **Section 3.2,:** What is the meaning of the numbers inside the maps? Are they the points to which the authors refer after? The forumulation of the standard deviation of the kriging leads to the linear relation pointed out by the authors, Have they used others definitions for the uncertainty estimation? For example, the one proposed by Yamamoto (Yamamoto 2000). **Section 3.3,:** Could be the underestimation of the zero-precipitation frequency considered a common problem of gridded datasets? **Section 3.4,:** ".. variance contraction is expected in any interpolation …" instead "is to be". Could be also "should be". The parameter considered, MAE, suffers less than the RMSE of the inflation due to low number of anomalous values so the

reasoning proposed would be more clear for that index instead the MAE. The figure shown only include one year and does not explain the relative error of 41% obtained for July. How this error is in comparison with others existing datasets?

References:

Herrera, S., Cardoso, R. M., Soares, P. M., Espírito-Santo, F., Viterbo, P., and Gutiérrez, J. M.: Iberia01: a new gridded dataset of daily precipitation and temperatures over Iberia, Earth Syst. Sci. Data, 11, 1947–1956, https://doi.org/10.5194/essd-11-1947-2019, 2019.

Peral C., Navascués B., and Ramos P., Serie de precipitación diaria en rejilla con fines climáticos Nota técnica 24 de AEMET.

P. Quintana-Seguí, C. Peral, M. Turco, M.C. Llasat, E. Martin, Meteorological analysis systems in north-east Spain. Validation of SAFRAN and SPAN, Journal of Environmental Informatics, 27 (2) 116-130, 2016.DOI: 10.3808/jei.201600335.

Quintana-Seguí, P., Turco, M., Herrera, S., & Miguez-Macho, G. (2017). Validation of a new SAFRAN-based gridded precipitation product for Spain and comparisons to Spain02 and ERA-Interim. Hydrology and Earth System Sciences, 21(4), 2187–2201. https://doi.org/10.5194/hess-21-2187-2017

Yamamoto, J.K. An Alternative Measure of the Reliability of Ordinary Kriging Estimates. Mathematical Geology 32, 489–509 (2000). https://doi.org/10.1023/A:1007577916868