Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-349-AC1, 2016

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Validation of a new SAFRAN-based gridded precipitation product for Spain and comparisons to Spain02 and ERA-Interim" *by* P. Quintana-Seguí et al.

P. Quintana-Seguí et al.

pquintana@obsebre.es

Received and published: 1 December 2016

We thank the reviewer for the time spent revising this article and for his/her thoughtful comments, which will help to improve the data analysis and the focus of the article.

Hereafter we will use italics to cite the referee's text and we will use regular type for our own answers.

Revision of "Validation of a new SAFRAN-based gridded precipitation product for Spain and comparisons to Spain02 and ERA-Interim". This study compares three different daily precipitation datasets in the Peninsular Spain and the Balearic islands. The main purpose of the manuscript is to present a new SAFRAN dataset for the whole Spain

and to provide validation metrics. I think that the dataset described in this study is highly relevant given several hydrological applications. The availability of this dataset for scientific purposes is also a valuable output. Nevertheless although the database created is really sound I find that the manuscript could be improved considering two main issues:

i) to include the validation of other variables obtained by the SAFRAN analysis and ii)to improve the validation of the temporal variability of the data generated.

The reviewer raised two main issues. Our opinion on these issues is:

1. Instead of including the validation of other variables, we consider that it is better to focus the article on precipitation only. We will improve the text to highlight more clearly that the focus is entirely on this variable.
2. We agree that the validation of the temporal variability of the data can be improved.

Hereafter we develop the reasoning behind these two opinions and answer, point-by-point the reviewer's comments.

I include comments on these two issues below. I would recommend the acceptation of the manuscript subject to the major changes suggested below:

1. The first paragraph in the introduction referring to current active research projects is unnecessary. This is not usually explained in scientific publications.

The references to current active projects were introduced to contextualize our work and to underline that these types of datasets have become strategic for national and international climate projects and programs. We prefer to maintain the references for these reasons, but if the reviewer's and editor's point of view is that this does not belong in a scientific paper, we can modify the text and remove those references accordingly.

2. Page 3, 25, What is IB02?

IB02 is the union of PT02, a daily precipitation gridded dataset for Portugal, and Spain02-v2 resulting in a merged Iberian daily precipitation dataset. Both components, PT02 and Spain02-v2, were developed using the same methodology ensuring the continuity and spatial homogeneity of the resulting dataset (Belo-Pereira et al., 2011). This will be clarified in the text: "... Spain; Belo-Pereira et al. 2011, who compared IB02 - an Iberian daily precipitation dataset built by joining two methodologically equivalent gridded products for Portugal (PT02) and Spain (Spain02-v2) - to global datasets and found that the global products produce better results in Western Iberia than on the Mediterranean side; ..."

3. Page 3, 29-30. The authors indicate that they are presenting a new SAFRAN dataset for the whole Spain. Nevertheless the other objective is less clear. If the authors are using this manuscript to present a whole SAFRAN dataset for the whole Spain including different variables, Why are they only focusing on the validation of precipitation and including a comparison with other precipitation datasets? I would find much more useful to present the new created dataset for the whole Spain as they do for section 3.1, providing more details and then to validate the different variables using observations of the different variables (e.g., wind speed, solar radiation, etc.). This would give much more consistence to the presented research and developed dataset instead of focusing only on the daily precipitation outputs.

We agree with the referee that SAFRAN should be validated for all its parameters/variables. However, the present work builds on the deep analysis and validation already described in Quintana-Seguí et al. (2008), Vidal et al. (2010) and Quintana-Seguí et al. (2016), where such general validation is performed. Thus, a new general validation would not add much new information with respect to previous work. As a consequence, we decided to focus the analysis on a single parameter, precipitation, and to make a more in deep validation of the dataset due to the relevant role of this variable in hydrology studies.

In order to improve the text, we can de-emphasize, in the introduction, the explanations related to other variables, focusing solely on precipitation, in coherence with the rest of the document.

4. Section 2. Spatial and temporal variability of precipitation in Spain is very complex, and if authors want to frame the developed dataset on the precipitation characteristics of Spain, they should describe in more depth this complexity. On the contrary, I suggest removing this section; it is not really necessary.

We agree with the reviewer that this section must be improved or removed, as now it does not add a lot of information on the complexity of precipitation in Spain.

5. Section 3.1: the SAFRAN meteorological analysis system should be described in more depth since this is the basis of subsequent analysis. In particular, the Optimal Interpolation Algorithm should be better explained. Why is the reason of using the period 1979-2014?

More information about SAFRAN's algorithm could be included in the paper, but our point of view is that this algorithm has already been explained in previous studies such as Durand et al. (1993, 1999), Quintana-Seguí et al (2008) and Quintana-Seguí et al. (2016a). Thus, it would be redundant to reproduce the same explanations in this article.

The period of study, 1979/80-2013/14, depends on data availability. As it is reflected in the manuscript, for all variables, except precipitation, SAFRAN depends on the first guess. The ERA-Interim dataset, our first-guess, starts in 1979, which defines the start date of our period. The end-period of our dataset is defined by the availability of AEMET data. As the data request to AEMET was sent in the second half of 2014, our data series finishes in mid 2014. Thus, the period considered was 1979/80-2013/14. We can rewrite the corresponding section in the text in order to clarify this point.

6. How are the climatological zones defined here?. I consider this is a key issue that

should be described in more depth.

Climatological zones are an important part of the SAFRAN approach. They should be areas of about 1000 km2 with weak horizontal gradients of the different variables, even though in practice, it is impossible to create zones that perfectly fulfil these requirements.

When we decided to implement SAFRAN in NE Spain we started our tests using river basin limits and AEMET's meteorological alert zones (Quintana-Seguí et al. 2016a). Although the difference was small, better results were obtained when using meteorological warning zones.

When we decided to expand SAFRAN to the whole of mainland Spain and the Balearic islands, we found that, in some regions, the meteorological alert zones of AEMET were too big; thus, we decided to subdivide them. We manually modified these large zones with the aid of a map of river basins and our own expert knowledge. In some areas it was very easy to define limits, just using basin boundaries; however in others, such as flat regions where there are no obvious discontinuities in the values of meteorological variables, the divisions were somewhat arbitrary. Note that, in this particular case, this is not a problem since the horizontal gradients are weak.

To sum up, the method used to define the map of zones shown in Figure 1, combined our own expert knowledge on the local climate, meteorological alert zones, river basin boundaries and the knowledge acquired in our previous study showing that the sensitivity to the limits of the zones is low.

We can include a description of the process in the manuscript.

7. Section 3.4. Authors indicate that the SAFRAN dataset was created by two different projects considering different time-spans. Is this approach having some impact on the temporal and spatial homogeneity of the dataset? The approach of considering different dependent and independent station data for validation according to the period

of analysis is complex and confuse. Authors should clarify issues of spatio-temporal homogeneity among the two projects.

We acknowledge that this is one of the weak points of this study, but, as we will explain, the impact on the results is very low or non-existent.

Ideally, we would have a high density of stations, with no data gaps and homogeneous time series for the whole period; however, in practice, this is not possible and thus we must find the right trade-off between data quantity, quality and data homogeneity. A very homogenous dataset would use only stations that have minimal gaps in previously homogenized time series. This would certainly ensure homogeneity but result in a low resolution dataset, given the sparsity of high quality data. If we want to include more stations to improve the spatial coverage and thus the spatial quality of the product, then stations with gaps must be used, therefore decreasing temporal homogeneity while improving spatial quality. There is no solution for this problem. One must make a choice.

SAFRAN favours spatial coverage over temporal homogeneity. This is the case of our own application of SAFRAN and also of the original French SAFRAN dataset. In Vidal et al. (2010), who made some comparisons between trends obtained using SAFRAN and homogenized time series, the quantitative differences between SAFRAN trends and the trends of the time series are apparent.

For a given time of analysis (every day for precipitation), SAFRAN ingests all available observations. It does an automatic quality assessment, comparing observations with analyzed values, but the data are not previously homogenized. As a consequence, the number of stations ingested is not the same, day to day. If a station does not have data for a given sub-period, it will not be used until it has data again.

In order to validate SAFRAN with independent data, we set some stations aside, randomly selected and making sure that there is a good spatial coverage. The selection is done automatically by a script at the beginning of the process, and these stations are

never ingested by SAFRAN. As our dataset was created in two different projects, this script was run separately in both projects and thus the stations set aside are not the same.

Regarding the impact of this break in the creation of the dataset on its quality and on this specific study in particular we note that:

1. Concerning the homogeneity of the dataset the impact is very low, because, by construction, the SAFRAN product is already not designed for homogeneity and the number of validation stations is low compared with the number of ingested stations (249 vs 988).
2. Concerning the quality of the validation, we already made the necessary steps in our methodology to avoid any problems. Thus, when we compare the dataset with independent stations we only do so for the SAFRAN-1 subperiod (1995/96-2006/07), which ensure the complete independence of the validation dataset. See Page 6, Line 25. Furthermore, most of the comparisons have been performed with dependent data, because there is no common independent dataset for both SAFRAN and Spain02, as these products were created in different periods of time and by different groups without any coordination (Page 6, lines 20-23).

Furthermore, Table 2 and section 4.1 show that SAFRAN is very robust, that is, the scores obtained when validating with independent stations are very close to those obtained when comparing to dependent data. This is also shown in previous studies.

To sum up, we think that, as SAFRAN is not optimized for temporal homogeneity, the impact of the break in the generation of the dataset is not really relevant. Furthermore, the number of validation stations is low compared with the number of stations ingested by the analysis, which implies that the homogeneity impacts should be low. The results of the validation are unaffected because our methodology did take this problem into account.

8. Authors should include more error/accuracy metrics to assess the performance of the gridded data. For example, the correlation coefficient is a very poor measure for temporal agreement between observed and modelled data. I recommend to have a look into the hydroGOF R package.

We can include some modifications to the manuscript in order to explore other metrics, being careful not to introduce redundant information. Thus, for example, we could study the possibility of using the Spearman correlation (rank based), in order to better tackle the effect of having many days of zero precipitation, which may dominate the Pearson correlation coefficient.

9. It is not possible for me to read the figure 3. Color scale is not very fortunate. The use of different length for the circles as a function of the values of the validation metrics would be a solution but it would be really useful to show a scatterplot with the metric values between datasets or maybe a boxplot showing the error metrics in the different datasets.

We can consider alternative colour scales or plot types, if the editor deems it necessary, but removing the black border line of the dots would probably be enough in order to make the figure clearer. Furthermore, scatter plots of the scores comparing ERA-Int and SAFRAN with Spain02 can be done too. Finally, it is also possible to add a boxplot with the spatial distribution of the errors.

A simple average of the correlation coefficients is not very suitable metric to have an idea of the average accuracy of the datasets. Temporal validation would gain if not only the agreement for the entire data is analysed but also temporal agreement for low/high precipitation days, dry spells but also considering possible seasonal influences. For example, it would be useful to know the temporal consistency of the datasets for the different months of the year to determine if the consistency is temporally different between the dry and humid seasons.

We didn't include this kind of detailed analysis because we wanted to keep the ar-

ticle relatively short, but we could add some more results building on the reviewer's suggestions.

Also the assessment of the accuracy/error metrics for different elevations would be useful to assess the potential applicability of the data. In this case, the division between stations located above and below 1000 m would be insufficient. Although the number of stations above a certain elevation is low (this is already stated by the authors) it would be very useful to assess the goodness of the prediction in these stations (some of them above 1800 m a.s.l.). Mountain areas are the principal water towers in Spain and where the main floods are generated. For this reason I consider extremely relevant to assess (even using the lowdata available) the goodness of the SAFRAN outputs in these regions.

We can add another altitude threshold on our analysis of the results pertaining to altitude.

10. Really I would focus in more depth on the validation of the temporal variability of precipitation than on the spatial variability of the average conditions. Usually modelled precipitation tends to reproduce well the average spatial precipitation patterns and the general precipitation seasonality. Thus, given the potential applicability of the SAFRAN dataset to force LSMs, the assessment of the temporal accuracy of the data is much more relevant than the spatial accuracy at the Spanish spatial scale.

As we said in a previous answer, we will explore the possibility of including one or more new scores, in order to improve this aspect of our study.

11. I find that the discussion section should be improved in more depth including limitations and potentials for the applicability of the SAFRAN dataset considering the proposed analysis related to the temporal precipitation accuracy. Given the potential applicability of the SAFRAN dataset for hydrological applications I find this much more relevant than assessing comparability with the Spain 02 dataset.

We can improve the discussion based on the comments of the referees and the analysis proposed in this document.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-349, 2016.