Hydrol. Earth Syst. Sci. Discuss., 9, C2972–C2978, 2012
www.hydrol-earth-syst-sci-discuss.net/9/C2972/2012/

Hydrology and Earth System Sciences Discussions

Interactive comment on "High resolution reconstruction of monthly precipitation of Iberian Peninsula using circulation weather types" *by* N. Cortesi et al.

N. Cortesi et al.

jcgh@unizar.es

Received and published: 12 July 2012

Many thanks for the comments and suggestions. We believe that they help to clarify the reading of our manuscript.

1) How is possible to obtain the maps in Figure 3 using only 16 grid points (Figure 2)?.

Response.

We believe that this issue results mainly from an interpretation of the data used to obtain Figures 2 and 3. In fact Figure 2 merely indicates the 16 grid-points used to com-

pute the geostrophic flow and vorticity indices (that are used to derive all the Weather Types). On the contrary, Figure 3a and 3 b use the entire SLP data (5° long x 5 lat°) available from the EMULATE database in the study area being interpolation of SLP during the day identified as any Weather Types. This interpolation was made using the mean of the SLP in those days using the EMULATE database.

2) How do you group the 26 daily WTs into monthly WTs? Averaging? The 26 daily WTs are valid for monthly time-scale.

Response.

We have used the usual procedure that most authors use when analyzing the impact of WTs on precipitation (e.g. Trigo and DaCamara 2000; Paredes et al., 2006) or electrical discharges (e.g. Ramos et al., 2011). Thus, the WT at monthly scale was calculated as frequency, i.e. summing the daily frequencies. Therefore a change in the temporal scale (using monthly or even seasonal averages) does not alter the number of WT (26) identified, neither their pattern. This reflects the fact that we only take into account the monthly frequency of each WT. In summary, each WT enter in the model as monthly frequency (individual year) from 1948 to 2003.

3) According the authors (page 6945, lines 11-12) 'multicollinearity does not reduce the predictive power of the model, i.e.: it only affects calculations regarding the independent variables and their coefficients'. I'm not agree. In fact, the multicollinearity makes difficult to diagnose the factors that are most important in specifying the predictand variable (von Storch and Zwiers, 1999). In addition, the sampling distributions of the estimated regression coefficients can become very broad, with the consequence that a regression equation may perform very badly when implemented on data independent of the training sample (Wilks, 1995). In general terms, the multicollinearity introduces redundant information and weakens the analysis. If you wish to avoid this problem, you must use the principal component analysis in characterizing the predictor variable (SLP field). At seasonal time-scale this procedure was followed by Muñoz-Díaz and

Rodrigo (2006).

Response.

Firstly, and to avoid confusion, we decided to delete the sentence in question (page 6945, lines 11-12) "... we did not remove any correlated WTs because correlations were not high 244 because multicollinearity doesn't reduce the predictive power or reliability of the model as a whole, i.e.: it only affects calculations regarding the independent variables and their coefficients".

Although we recognize the problems raised by the reviewer, we firmly believe that to a large extent the problem is almost non-existent due to the simultaneous application of the stepwise-regression coupled with the comprehensive use of a very stringent cross-validation procedure. Please note that problems related to over fitting are highly mitigated by adopting a leave-one-out scheme for cross- validation (Wilks 1995), i.e. by successively using a single observation from the original sample for validation, and the remaining observations as the training data. This aspect will be further emphasized in the new version of the manuscript, and here we present the main reason that suggests us to work as we did. As the referee suggests, from a mathematical point of view it's much better to perform a PCA prior to the regression to avoid multicollinearity. In fact we tried this approach with another model based on a principal component regression analysis (PCRA), but there is a practical limit of the PCRA model: the predictor variables obtained from the monthly WTs frequencies can also assume negative values, while in a Multi-Linear Regression (MLR) model without PCA the monthly WTs frequencies are always positive or null. In that sense we can choose to force the regression coefficients to be always non-negative too (as we did). On the other hand, if we applied PCA and discard any WT unfortunately it greatly affects the outcome of the model because WTs are defined globally for all IP series, so when two predictor WTs are significantly correlated and one is removed, it is removed from all the remaining MLR models, so its contribution to the total monthly precipitation disappears entirely all over the IP, not only in a few series. For example, the NW and the A.NW (Anticy-

clonic Northwesterly) weather types during January are highly correlated (r=0.64). If we choose to remove one of the two WTs from the regression, the A.NW for example, its contribute to total January precipitation disappears (even if it is very influent) and the regression coefficient of the NW weather type increases at all sites, to account for the missing contribute of the A.NW, artificially augmenting the spatial distribution of the mean rainfall amount due to the NW. To conclude, to minimize multicollinearity without removing a priori any WTs we stopped the stepwise regression when the new predictor WT variable added failed to improve the local RMSE of the selected series and month of a fixed threshold. In this way, any correlated WT may be excluded from the regression for some series if they don't improve the RMSE very much, but may be included in the regression for other series if they improve sensibly the RMSE, even if we know that they are highly correlated. We are conscious that this solution doesn't remove completely multicollinearity from the series where and when two or more high-correlated predictor WTs are picked up by the MLR, but it was the best possible compromise to extract all the information from WT. We would gladly accept any suggestions of the referees about this matter. Therefore, we are confident that the problems raised by the reviewer: "...with the consequence that a regression equation may perform very badly when implemented on data independent of the training sample (Wilks, 1995)" are virtually absent.

4) What are the columns A and B of Table 3? (page 6948 lines2-3).

Response.

We are sorry. An error was produced. The paragraph, referring comments to Table 3, should be as follows.

“We present in Table 3 the monthly percentage estimation of precipitation (in mm) by WTs over the total observed monthly precipitation. Please note that columns do not sum 100 because we did not present the constant term, and because the value are calculated dividing the amount (mm) of precipitation predicted by the amount (mm) of precipitation observed, and the result can be different to 100%."

5) Please, specify the explained variance of the model. It is a very intuitive parameter, which allows check the goodness of the model and study if it is necessary to enlarge the analysis, looking for other predictors to explain the behavior of the predictand.

Response.

In Figure 4 it is shown the "r" Pearson correlation between predicted and observed precipitation pixel by pixel. The maps of the explained variance (r2) of the model show similar patterns to figure 4, and given that our focus was on the spatial analyses, i.e. subregional differences, gradients etc, we did not included r2 maps, because of disparity of model performance across IP it makes no sense to rely on a global value for the entire IP. This is the reason because we preferred to present instead of global variance, spatial differences of correlation.

Notwithstanding we agree with referee about the most "intuitive " way to show the goodness of model and the availability for comparison, so we suggest to change the original Fig 4 (r Pearson) by new Fig. 4 Monthly precipitation explained variance (r2) from October to March by the regression model (Comments along the text should be checked).

6) Have you analyzed other factors in addition to SLP? In particular, for the Mediterranean fringe, where the model is weaker, it would be interesting to consider SST and soil moisture as predictor variables. A brief discussion on the possible role of these factors and the contribution of convective rainfall would be necessary to complete the study.

Response.

Not at present. This study is a starting point of WT and monthly precipitation in the IP focused on detect the global scheme, the main WT contribution across IP, the spatial differences, the gradients, etc. The ideas suggested by referee are interesting, and we

would be able to include in the final manuscript in the following sense: "The general results of the study reflect with high accuracy a very well defined division of IP accordingly the WT analysed, that were detected primarily at seasonal scale in Muñoz and Rodrigo (2005), i.e. the northern coastland, the Mediterranean fringe including Ebro basin and the central west areas, with different response to WT approach. Particularly in the eastland areas the effects of SST of Mediterranean Sea, and also the increment of convective processes, with very local effects, could be one of the main reason of the lower accuracy of the WT approach presented here".

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 6935, 2012.

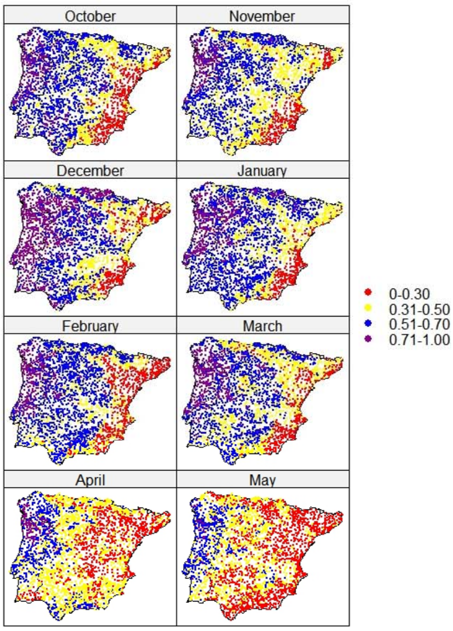
October
November
December
January
February
March
April
May
0-0.30
0.31-0.50
0.51-0.70
0.71-1.00

Fig. 1.