Dear Editor

First of all we would like to acknowledge the thorough revision made by Ref. 2 and his/her comments that helped us to clarify and improve the manuscript. Given that Ref-1 suggestions were included in the main text that was modified substantially, some of the comments raised by Ref-2 were already considered in that new version of the manuscript.

We made a serious effort to cover systematically all comments raised by Ref-2 because they are written in extended paragraphs,.

General comments.

The reviewer has focused mainly on the following questions:

a) use of the evaluation parameter CV (name and concept),
b) constraining effects to positive values,
c) insufficient explanation of the use of CWTs in the predictor,
d) missing interpretation of the constant term in the model
e) verification of the Gaussian assumption.

Additionally, the reviewer suggests to delete the section where we established the relationship between river discharge and WTs

All these different points are expanded as specific comment and are dealt separately on a point-by-point basis and where we try to answer carefully to each one.

Specific Comments

Ref´s comment	Our response
2.1 Title. Is the reconstruction the main topic of the paper? “High resolution” reminds me on a grid but the approach is station-based and for the approach, it does not matter if you have three or three thousand stations. Reading the paper, I would have entitled it something like “Modelling monthly precipitation sums with circulation weather types for a dense network of stations ...”	Perhaps there has been confusion and the initial title is not the best option. Our approach is to model the relationship between WTs, as defined, and the highest density of monthly precipitation stations available for Iberia Peninsula, and given the results to show some applications, i.e. feasibility for reconstruct long term precipitation series. So reconstruction is not the main topic of the paper while of course the model validation implicates the reconstruction of 3030 series of monthly precipitation for study period (see cross-validation procedure) Accordingly, we have adapted the new title suggested by the reviewer as follows: **Modelling monthly precipitation with circulation weather types for a dense network of stations over Iberia.**

Abstract.

We have rewritten completely the abstract following the main topic indicated below. Given some specific comment we try to answer each one individually (being also useful to check the new version of the abstract)

Ref. Comments	Our response
giving the number of stations is not really a “resolution”	We agree with ref´s about the word resolution, which is not the most appropriate in that context. In general we replace it by ***high density***
“stepwise linear regression model with forward selection” does not well specify the model. The crucial information needed is the Gaussian assumption for the response (monthly precipitation sums) and the frequency of CWTs as predictors, as well as the independent treatment of stations and month.	In a linear regression the assumption of following a Gaussian distribution is only required for the regression residuals, and not for the response variable or the predictor variable (please see Wilks 2006 par. 6.2.2.) Furthermore, models were developed independently for each station and month of the year. In any case, the abstract redaction has been modified, and also the presentation of model, please, see for details the item
The coefficient of variation (CV) is commonly defined as CV = σ/μ, the ratio of the standard deviation and the mean for a random variable. Without reading the rest of the paper and understand that you redefine this concept in your work using a relative error, I could not understand the abstract.	We have deleted all comments on the CV in the abstract. The confusion about CV used is due to the existence of slightly different definitions of the CV measure. The CV can be defined as the standard one quoted by the referee (that we never use) and the so-called “CV of the RMSE”, which is defined as the RMSE normalized by the mean of the observed values of the variable (monthly precipitation in our case). **We define more precisely n the text the CV used as CV of the root mean standard error, and change all the indication of CV by the CV_{RMSE}**
is it fair to promise a monthly precipitation field when having only station-based data?	We’re not sure about this criticism because one can derive a monthly field from gridded and station based information, as long as the number of stations is sufficient (as it is the case in our study). Our approach uses the highest spatial density of monthly precipitation for IP combining two dataset from Spain and Portugal during an extended period (1948-2003) with an overall density of 1 station /200 km^2.

Introduction

Ref´s comment	Our response
"This explains the generalized recommendation of high density precipitation which requires a database for regional analyses."	We have rewritten the sentence as follows: ***This explains the generalized recommendation of using high density precipitation database for regional analyses (Auer et al., 2005; Brunetti et al., 2006).***
The model selected was a forward stepwise linear regression derived from that of Trigo and DaCamara (2000) and of Goodess and Jones (2002)." Forward selection is the approach you followed, right? Standard linear regression with CWTs as the predictors is the model, if you wish.	We agree with the reviewer that this sentence was misleading. Therefore we have deleted these sentences (see comments below in Methods)
The discrepancy between daily CWTs and monthly precipitation is not resolved in the introduction which is probably puzzling for many readers at this stage.	CWTs are computed at the daily scale (monthly CWTs would not be very useful). However, to establish their link with precipitation (at the available temporal scale, i.e. monthly) one must use the temporal resolution of the precipitation dataset (monthly). Therefore we computed the monthly occurrence of the daily WTs and only afterwards we attempt to relate with the precipitation at each station. We try to explain better our approach at the end of introduction rewriting the objectives as follows: ***The research is conducted not on precipitation variability itself, but on the nature of its variability. Accordingly, the objective of this paper is twofold: firstly modelling the relationship between WTs and precipitation with the highest spatial detail available at present in the IP at monthly scale; and secondly to show the usefulness of such an approach in long term precipitation reconstruction. This paper is the starting step to provide the opportunity of extending the reconstruction of monthly precipitation for a very high density of stations as far back in time as 1850, because catalogues of circulation weather types are now available since then. These reconstructions at high spatial detail would provide in the near future the long-term contextual framework of precipitation variability and trends in the IP, thus allowing considering the recent changes of precipitation monthly***

	distribution within a more global context from the middle of the 19th century. Therefore, while the main effort of this paper is focused on the evaluation of models performance during the 1948-2003 period, we will also assess the potential of this modelling approach by applying the validation with three of the longest series of monthly precipitation available in the IP.
You might want to consider another paper modelling precipitation with the CWTs, not for the IP, however [Maraun et al., Extremes 13:133-153 (2009)] or as an overview for statistical modelling of precipitation [Maraun et al., Rev. Geophys. 48:RG3003 (2010)].	Thanks for providing these references. We have included the reference to the study linking WTs and extreme precipitation in the UK (Maraun et al., 2010) as it fits well our revision of precedent research.

Methods

<table>
<tr><td>Ref´s comment</td><td>Our comments</td></tr>
<tr><td colspan="2" align="center">2.4 Precipitation data</td></tr>
<tr><td>Some information on the gaps in the series would be helpful, or if gaps have been filled this would also be interesting to know.</td><td>The database is composed only by complete monthly series after an exhaustive quality control and reconstruction processes using high quality reference series explained in detail in Gonzalez-Hidalgo et al. (2011). Thus, all the precipitation series considered here have no monthly gaps during the period analyzed. We agree that this was probably not clear in the previous version of the manuscript, therefore we have rewritten the sentence as follows:

The series come from an exhaustive quality control of original information and reconstruction processes; thus all of the time series used in this work are complete (no gaps), and free of anomalous data and inhomogeneities (details can be found in Gonzalez-Hidalgo et al., 2011, and Lorenzo et al. 2011).</td></tr>
<tr><td colspan="2" align="center">2.5 Circulation weather type classification</td></tr>
<tr><td>If one does not know the approach used here, it is not easy to understand. Instead of pointing to a plethora of literature using the CWTs, you might want to spend more time on explaining them. A set of equations and some classification criteria (exemplary if you wish) additionally to Fig.2 could be very helpful here.

Furthermore the occurrence of the individual CWTs for the individual months as a bar plot could be helpful. You promise to give similar information in Tab. 3, where I did not find it.</td><td>We agree with the reviewer and have therefore included some additional information on the computation of the WTs based on the original work of Trigo and Dacamara (2000).

In relation to Table 3 we are sorry for the contradictory information given in in the original version. As explained before (when answering to ref_1 a month ago), the text referring to this Table has been changed (see comments to ref-1) and added a new paragraph in text.</td></tr>
<tr><td colspan="2" align="center">2.6 The model</td></tr>
<tr><td>Even if it is seemingly simple, you should explicitly write down the model as a regression equation, mention that the response is the expectation of a Gaussian random variable, namely the monthly precipitation sums and explain what is actually used as predictor</td><td>The section has been completely rewritten to ensure maximum clarity. Among several other points it should be emphasized:

The regression used is a standard multiple regression (y = a + b1x1+b2x2….bnXn) with forward selection as indicated in text.

We agree that it can be useful to provide an example of the regression equation obtain. Therefore we introduce as an example the</td></tr>
</table>

	general equation
I suppose it is the days of occurrence of the certain CWTs relative to the length of the month). This makes clear that the predictors are dependent, they should sum up to one (or total days in the month).	We believe that we have not been so clear so we suggest a new paragraph presenting and explaining how the variables enter in equation as follows: ***The model selected was a multiple linear regression (with a stepwise forward selection procedure) adapted from that of Trigo and DaCamara (2000) and of Goodess and Jones (2002). It considers the occurrence at monthly scale of all 26 WTs as predictor variables, and the corresponding monthly rainfall totals as the predicted variable along the study period 1948-2003 (Eq. 1).*** Also we present the general structure of model and discussion about the selection procedure (see comments ref-1)
You mention "The forward selection criteria". To my understanding, forward selection is an approach to predictor selection, i.e. choosing those factors which are relevant for describing the response (precipitation). I would NOT call this a "criterion". In order to perform this selection, you do however need a criterion to measure improvement of the model due to addition of a new predictor; generic criteria are Information criteria as the AIC, BIC, etc. or a likelihood-ratio test or even a cross-validation experiment. All these criteria somehow involve the model complexity, i.e. the number of parameters (predictors) used. The idea is to find an optimal trade-off between "model error" and model complexity (Occam's Razor). The before mentioned criteria do this.	We agree with the reviewer that the notion of criterion (criteria) was misleading in this context. Therefore we have removed the word "criteria", substituting this by the word "procedure".
I do not see how your criterion the "CV" does that since it is not clear what a significant improvement of the model is. Your ad hoc value of 0.01 seems arbitrary to me and without any theoretical foundation. I browsed through the new edition of Daniel Wilks text book (I think you referred to a previous edition) and could not find a hint to your criterion, please cite the chapter you are referring to. A short explanation on why you are using this measure of relative error would be helpful. If it is only to make the RMSE comparable between	As stated above we have used the so-called "CV of the RMSE", which is defined as the RMSE normalized by the mean of the observed values of the variable (monthly precipitation in our case). It is the same concept as the first definition of CV, except that the RMSE replaces the standard deviation. To avoid confusion we replaced in the text the word CV with the more accurate CV_{RMSE} Please, see Wilks (2006), Chapter 6.4.3

stations, this is already a valid reason.	("Stopping Rules"), where it states that "The stopping criterion can also be based on the MSE (…)". The CV (RMSE) criterion we adopted is really the same MSE criterion of Wilks called with a different name: as the referee inferred, it was used only to make the MSE (or the RMSE) comparable between series by normalizing the RMSE for the mean observed precipitation (remember that the RMSE is the square root of the MSE). We needed to introduce a threshold of 0.01 because the CV calculated by the leave-one-out cross validation and applied to a single time series and month not always reaches a minimum value for some combination of K predictors and starts increasing thereafter (as shown in dashed line of figure 6.18 of Wilks text book for K=3): sometimes the CV remains almost constant or decreases slightly when the number of predictors increases beyond a certain number. In this case we have to lower the number of predictors because we are not only interested in getting the minimum CV value, but even in minimizing the number of predictors to avoid over fitting, and in this case it is not convenient to choose the maximum number of predictors only because the CV decrease slightly in comparison, for example, to half the number of predictors. The threshold of 0.01 assures that the number of predictors is always balanced to take into account all possible cases.
To my understanding, you use a constant offset for each month, thus there is a seasonal signal modelled by this constant, I think this interpretation is worth mentioning. Another efficient approach could be to consider only monthly precipitation sum anoma-lies and combine the data of three or four month to one season.	Our study is focused at the monthly time scale because Iberian precipitation characteristics change significantly on annual scale but also within each season (namely in spring and autumn) therefore requiring a modelling methodology at the monthly scale We have also developed models based on anomalies; overall it worked fine, but presented a lower performance skill than the corresponding models based on absolute values of monthly precipitation and non-negative parameters
What about the residuals? Are they sufficiently close to the Gaussian assumption made? A plot (e.g. QQ-plot) would be nice to illustrate that. The distribution of precipitation, at least on a daily scale, is usually skewed and	We have rewritten the sentences on this issue as follows: We used the coefficient of variation CV_{RMSE} (defined by equations 2 and 3), and the explained variance measured by r2,

assumptions other than Gaussian are used, e.g. Gamma for a daily scale [Ambrosino et al., J. Climate 24:4600-4617 (2011)].	between the observed and the modelled precipitation, as indicators of goodness of model. Concerning the evaluation of error, the RMSE (and its CV_{RMSE}) is a good estimator of error even when the regression residuals don't follow a Gaussian distribution. In fact even if a serie shows a low RMSE for some months this is sufficient to know that the model performance is good during those months, although it's not possible to calculate a confidence interval. Furthermore, it allowed us to compare values obtained for different stations. This is better than what is achievable with other measures, such as the root mean squared error or the mean absolute error, as these evaluation parameters can change dramatically between stations.
Leave-one-out cross-validation is not the same as jack-knifing, the latter is commonly used to estimate the bias of an estimator, while the former is used for model validation (cf., e.g. D. Wilks textbook) .	We agree with reviewer's comment and delete the word "jack-knife" from the article and left only leave-one-out cross validation instead.
"measure model validation" sounds strange to me. In the context of model validation one can measure the model performance, fitness, quality, etc.	Again we agree and have deleted the word
2.7 Model validation	
It is not clear to me how the cross validation experiment was performed. Please add a sentence explaining that briefly. Something like: for each station and each month a one-out cross-validation experiment was carried out. One data point (i.e. one month) has been separated, the model parameters are estimated for the remaining data points and the model performance for the data point left out has been calculated using ... This procedure is repeated until all data points have been left out once.	Following the reviewer's suggestion we have introduced the following sentence: ***Model validation was performed by means of a leave-one-out cross validation over the regression period for all 3030 series of monthly precipitation of IP. Specifically, for each station and each month the experiment was carried out as follows: one monthly precipitation series were excluded, and then we estimated the model parameters for the remaining data points and the predicted precipitation for the series discarded was then calculated.***
An interesting piece of information would be the bias of the model, since the linear correlation coefficient does not report this. It is, however, included in the RMSE and thus in your "CV". Significance for the correlation coefficient should be reported.	As was written in the draft, we did not include the graph of the global bias because it did not show any relevant pattern. We indicate such point in the original draft (see below further comments) We substitute the Pearson correlation map

	for explained variance[2]
Recall that you model the mean of the response variable and you expect a dispersion around this value (residual standard deviation). This dispersion should be reported, at least in Fig. 9 as error bars.	We agree with the reviewer that some measure of dispersion is necessary to be included. However, the inclusion of error bars in figure 9 would difficult its reading. Therefore we decided to include a new Figure 10 in which we show the dispersion of observed vs predicted precipitation for the long term reconstruction (Lisbon, Madrid and Valencia) for the January precipitation during the 1948-2003 period.
You mention that the residuals are “normally distributed around the null value” but you also say that “the width of the left half is usually twice the width of the right half”. This is contradictory. If the last statement is true, your modelling assumption is not even approximately fulfilled.	We agree with the reviewer's criticism that these 2 messages can be contradictory. In fact the text had to be rephrased better because, as stated before, not all the residual series follow a normal distribution.
You refer to column A in Tab. 3 which is missing. Column B (not indicated as such) shows “the percentage estimation of precipitation by WTs over the total observed monthly precipitation.” A more consistent approach in my eyes would be to divide by the total predicted precipitation to obtain the contribution of a WT.	Please, see comment to ref-1. There was a mistake in the presentation of table 3.
2.8 An example of reconstruction of long term monthly precipitation in the IP	
As far as I understand, you “reconstruct” (hindcast) monthly precipitation for three stations as an example. I could not find further reconstructions presented in this manuscript. Reconsidering the title in this light, I find it not being well chosen. Furthermore, I would expect confidence biands/error bars for the “reconstruction”.	We agree with the reviewer criticism on this issue. Therefore we have changed the title in order to remove the word "reconstruction". See also answer to question 2.10. In relation to the error bars request please see the answer to the third issue of question 2.7. But this information in any case is given in tables
2.9 River flow modeling	
Just after the results of your main point and just before the discussion, there is a new chapter. It starts in the tone of an introduction which I found very irritating here. Since the river flow modelling part is not sufficiently described but only briefly, I recommend to take it out.	We agree with reviewer's comment and decided to remove this section that was only introduced to prove the concept that WTs could be used to model river flow at the monthly scale.
2.10 Discussion and conclusion	
Here, you start with: “The circulation weather type classification devised by Trigo and DaCamara (2000) has been successfully applied to reconstruct and	We agree with the reviewer that the use of the term "reconstruction" when applied to the calibration/validation period (1948-2003) can be misleading.

validate monthly precipitation for the 56 yr of the period 1948–2003 at 3030 Iberian site locations ...”. As mentioned earlier, I found the reconstruction for only 3 sites in the paper.	Thus we have agreed to change the title of the paper (see answer to first comment) and also to de-emphasized this aspect as the reconstruction could have been applied to more stations other than the 3 long-term series mentioned (Lisbon, Madrid, Valencia).Therefore we have rephrased this sentence accordingly.

Tables and Figures

Ref´s comment	Our response
Some results presented in the form of tables could be as well presented as figures (e.g. bar plots) which I consider as preferable. For example Tab. 3 and the missing "Column A".	See previous comment about the error produced in table 3
How is the spatial interpolation realized in Figs. 6 and 7? Why is that used here? You could have presented it in the same way as Figs. 4 and 5. I don't see added value.	All the maps presenting different aspects of model (Figs 4-8) were produced originally using point data results (i.e. one value for each observatory). The main reason for presenting Figs. 6 and 7 using interpolation (ordinary kriging) was to improve their the clarity . We are able to present these results in both formats (i.e. based on stations or interpolated) to decide the final version as Editor request. Here, in order to avoid confusion we present the original maps with all individual stations represented.
Figs. 9 and 10 have too small labels and there are no confidence bands. Helpful for a validation would be a scatterplot (predicted vs observed).	We have removed the original Fig.10 after accepting the reviewer's suggestion to drop the entire river flow section. Concerning Fig.9 we believe that error bars for each year would make the reading particularly unappealing. Nevertheless we agree that some visual information on the errors should be available to readers. Therefore, besides the information provided in Table 3 we have obtained the scatter plots relative to the comparison between observed and modeled precipitation (NEW FIG. 10,)
Label text is too small in Figs. 3a and 3b.	We are not sure of the problem. The original resolution was sufficient (considerably better than what we see here) Perhaps this is a problem of editing