

Interactive comment on “Weather model performance on extreme rainfall events simulation’s over Western Iberian Peninsula” by S. C. Pereira et al.

S. C. Pereira et al.

susana.cardoso@ua.pt

Received and published: 30 November 2012

Dear Dr Morin,

We thank the editor and the referees for their thoughtful comments and critiques. We acknowledge the time and effort that editor and reviewers had put into assessing the previous manuscript. We have adopted all suggestions, including rewriting the introduction, the methods, the results and the discussion sections and clarifying the ability of the model in reproducing this precipitation event. Also, as suggested by both referees, the title was changed and some new analysis was performed. This entire

C5580

process resulted in a completely new manuscript but based on the previous version. Nevertheless, since the backbone of the manuscript remains the same, we can still address the major concerns issued by the editor. Thus, all the previous calculations were revised and a new analysis was performed with new statistical parameters. More, as suggested, the rain-gauge stations were split into multiple groups according to elevation; this lead to clarification of the results and conclusions. Specially, the agreement between the model and observations; the title was changed and the use of "extreme episode" was changed to meet the precipitation characteristics of December, 2009. A month with a superior total amount of precipitation. The new title is "Simulation of a Persistent Medium-term Precipitation event over the Western Iberian Peninsula"; The hydrologic implications are now discussed; The organization, structure and the text were modified and most parts were rewritten. We think that the manuscript benefited with these revisions and we hope that will now be suitable for publication. The elaborated responses to the recommended changes are detailed in the following point-by-point responses to comments. We shall look forward to hearing from you at your earliest convenience, Yours sincerely, Susana Pereira

Editor Comments

editor 1. apparently not so good fit between the model and observations in contradiction to the conclusion about the capability of the model to reproduce the extreme event;

The manuscript was reviewed and some new analysis was included. In the new "Conclusions" section the authors expressed the disagreement between model data and observations. It is also mention that although, simulating precipitation does not appear to have sufficient quality for hydrological applications, were the corrected precipitation amount and timing are required, the periods of precipitation occurrence was reasonably simulated. The overall model accuracy was similar for all altitude classes, for the three experiments: highest for lowlands and highlands. Precipitation simulated in areas located in rough terrain and deep valleys tend to be less accurate.

C5581

editor 2. the concept of “extreme” appears in the title but it is not well justified in the analyses, and, it is also not clear if the month of December 2009 included one or more extreme events or it has extreme monthly rain amount (the title suggest the first option), and does it really extreme in terms of return period,

The "Introduction" section was modified in a way that better explains the use of December 2009 as a case study. The use of “extreme episode” was considered within the scope of hydrological applications. We explored the occurrence of extreme events in more detail, by analysing the return period of the maximum daily rainfall in December. The results indicate that the rainfall had a high return period in the S-SE part of the study area, in part due to the occurrence of a high rainfall day in December 6. The maximum rain day for December 2009 had a return period between 7 and 18 years for this part of the study area, although the return period for the maximum 24 hour rainfall should be lower (c. 5 years) and may not be classified as an extreme. Also, this was probably not the result of events with steady rainfall during 12-24 hours. Regarding the concerns of both referees, suggesting a lower importance for intense rainfall events was therefore misleading (actually we meant to say that high rainfall was not the result of high-intensity short-duration events, but rather of episodes lasting for c. 24 hours, but the sentence was confusing). Section 2.1 was rewritten to include this analysis. The authors reinforce their definition of extreme event in the scope of hydrological applications. This phrase reinforces the idea that, individually, the events may not be extreme, but at the end of the month, the sum of the events is relevant for hydrological applications. The National Weather Service (I.M.) reported December 2009 as the highest month with the accumulated precipitation of the decade (I.M., 2010).

I.M., Instituto de Meteorologia, I.P., 2010: Boletim Climatológico Mensal – Dezembro 2009, Instituto de Meteorologia, Lisboa, available at www.meteo.pt.

editor 3. Implications to hydrology are weak and not discussed,

The implications for hydrology are now discussed in the conclusion section. A new

C5582

paragraph was added. Now, the conclusions section begin with the motivation behind this investigation and a new paragraph about the hydrological applications are mentioned; as well as the new results from aggregating the stations by altitude and the use of an additional set of verification measures. Now we mention the poor performance of the model in providing data for an event-based hydrological modelling, where the corrected precipitation amount and timing are usually required. Although the results are not promising, the model was able to give information about the precipitation, as can be deduced by the low mean error, and correctly identifies the temporal precipitation pattern. Also, the good model performance in simulating the spatial feature of precipitation provides an additional advantage for hydrological modelling, in particular for the complex orographic regions and considering watershed precipitation.

editor 4. the whole manuscript should be improved in terms of grammar and structure.

The manuscript was extensively reviewed. We acknowledge that the manuscript might have been confusing so several sections were modified. For both, figures and tables, the captions were changed to include more detailed information. The plot units, axes labelling, colored symbols and legends were included or rearranged. The mislabelling issues, grammar and references were reviewed and corrected. In this process, some old tables and figures were improved while others were discarded. Also, some new tables and figures were added to reflect the changes made in the manuscript.

Major Comments - Anonymous Referee #1

1. I question if this is enough to actually declare that the “extreme episode was successfully reproduced” (pg 9164, L18), Especially since the amount of precipitation is not reproduced well (pg 9164, L16 and pg 9171, L10-11).

We agree with the referee but the overestimation of precipitation amount is in line with the other results obtained for mountains regions by studies referenced in this work. Also, although the amount of simulated precipitation is not in agreement with the observed one, the temporal variation of the hourly precipitation was well reproduced. We

C5583

must also take into consideration that we simulated an orographic complex area with a high-resolution grid (1-km) where local phenomena may not be well described in the model due to misrepresentation of elevation or bad conjunction of physical parameterisations for this particular region. The authors included new text in the "Conclusion" section where they expressed the disagreement between model and measure data. 2. Additionally, the correlations between observed and modelled time series are low, as described at the end of section 3.1 but not shown (pg 9172, L1-6). The authors claim that it is due to timing and space issues (pg 9172, L5), even though they claim that the timing of precipitation is well represented by the model (pg 9164, L15-16). A point correlation for the time series at each observation station with the time series at each grid point for each of the model runs would tell you exactly how well each observation station correlated in space with the nearest grid points. Additionally, if you repeated the analysis done in Figure 3, looking at anomalies from observations rather than absolute quantities, you would be able to better assess the timing errors. The combination of these two results would allow you to tell which factor has a larger contribution to low correlations, the timing error or the spatial error.

This was a good point and was investigated as follows: All the correlation calculations were reviewed. In doing so it was found an error that completely transformed the meaning of the question. In fact, the correlations found are not low but right the opposite. This has been corrected in the manuscript. As suggested, a point correlation for the time series at each grid point was calculated and the results were included in the "Results and discussion" section with the appropriated figure (Fig. 2). For convenience, only the RunRef case is shown. We went deep in this question and the correlations between lag -24 and lag +24, for the observed and model series, were also calculated. In the results section a new table (Table 5-7) was include with the correlation values, for each of the model runs. The mean absolute deviation was calculated and a table (Table 4) with the respective analysis was introduced in the "Results and discussion" section. 3. Other analysis might include splitting the stations into multiple groups according to elevation. It appears that you credit much of the error signal in space to high

C5584

terrain stations not lining up appropriately with the model grid, yet the highest elevation stations (S14, S24) appear to have some of the smallest errors (combining information in Table 3 with Figure 2). The largest errors occur at stations at intermediate elevations.

As suggested by the referee, stations were grouped into five classes, according to altitude, and statistics were computed for each class. The division by altitude classes was made considering a 200 meters interval. This methodology helped in finding some differences between simulations or locations that are presented in the manuscript. Also, the methods were revised and new verification measurements were used. The "Methods" and the "Results and discussion" sections were modified in order to consider these modifications. Some old Tables (Table 1 and Table 2) and figures (Fig. 1, Fig.2 and Fig.3) were remade and new ones were introduced changing the order for both figures and tables. This new analysis revealed that, the overall model accuracy (RMSE) was similar for all altitude classes, for the three experiments: highest for lowlands and highlands. Precipitation simulated in areas located in rough terrain and deep valleys tend to be less accurate.

4. A final major concern with this study is the nature of the month of December 2009 being described as extreme. An overall extreme month does not equal a month containing extreme events as described in your title. The authors even mention that the high levels of rainfall for the month could be attributed to a greater number of rainfall days rather than intense events (pg 9167, L10-11).

The authors modified the text in order to exclude the expression "extreme". The introduction section was modified in a way that better explains the use of December 2009 as a case study. The use of "extreme episode", short-duration/high intensity or long-duration/low intensity, was considered within the scope of hydrological applications (see response to editor #2).

5. Finally, the conclusions section (really more of a summary section rather than conclusions) could do a much better job synthesizing the analysis and talking about the

C5585

implications of the results on applications for hydrologic models. The authors do not mention the motivation in the conclusions section, or anywhere else, after the original reference at the end of the Introduction section.

We acknowledge that the manuscript might have been confusing so the "Conclusion" section was modified according to the suggestions presented by the referee. Now, the conclusions section begin with the motivation behind this investigation and a new paragraph about the hydrological applications are mentioned; as well as the new results from aggregating the stations by altitude and the use of an additional set of verification measures.

Now we mention the poor performance of the model in providing data for an event-based hydrological modelling, where the corrected precipitation amount and timing are usually required. Although the results are not promising, the model was able to give information about the precipitation, as can be deduced by the low mean error, and correctly identifies the temporal precipitation pattern. Also, the good model performance in simulating the spatial feature of precipitation provides an additional advantage for hydrological modelling, in particular for the complex orographic regions and considering watershed precipitation. The overall model accuracy (RMSE) was similar for all altitude classes, for the three experiments: highest for lowlands and highlands. Precipitation simulated in areas located in rough terrain and deep valleys tend to be less accurate.

6. In addition to the concerns mentioned above, the manuscript contains several grammatical errors and technical problems, such as, a mislabelling of the last several figures, making it quite difficult to read and comprehend.

We understand the concerns of the reviewer. The manuscript has been completely revised and rewritten. In some sections, such as the "Introduction", the text has been reorganized adding new concepts more modelling oriented and rearranged some parts to make it more understandable. In addition to changes in the introduction, we have made the corresponding adjustments to the abstract. In the "Material and Methods"

C5586

section some parts underwent minor changes but others are totally new to reflect the inclusion of new methods. The results and discussion section and the conclusion section were rewritten to reflect the analysis with new verification measurements as well as the grouping of stations by altitude. In this process, all of the figures and tables have been reviewed. For both, figures and tables, the captions were changed to include more detailed information. The plot units, axes labelling, colored symbols and legends were included or rearranged. The mislabelling issues, grammar and references were reviewed and corrected. In this process, some old tables and figures were improved while others were discarded. Also, some new tables and figures were added to reflect the changes made in the manuscript. Specifics Issues - Anonymous Referee #1

7. The title is misleading. This study does not evaluate model performance on extreme events. It evaluates the models ability to reproduce observed precipitation fields over the course of a single month, a month in which higher than average precipitation fell, not necessarily as few extreme events but probably as many moderate events.

The title has been changed to Simulation of a Persistent Medium-term Precipitation event over Western Iberian Peninsula. The title was chosen to meet the nature of the event (see response to #6).

8. (pg 9164, L12) Simulation 2 is referred to as data assimilation but it is not mentioned this way again in the entire manuscript. It was initially confusing to me and took me some time to realize that it was just a different kind of nudging from experiment3, since L13 describes simulation 3 as nudging.

Section 2.3 was rewritten, having a more clear definition of the numerical experiments. The data assimilation term was replaced by observational nudging.

9. (pg 9166, L23) 2300mm must actually be 2300 meters.

Actually, mean annual rainfall in the study area can reach as high as 2300 mm/year in the highest altitudes (less than 1000 meters). This affirmation is based on the fact

C5587

that the mean annual rainfall for the 1961-1990 normal at the Caramulo rain gauge, the highest with long-term data in the study area (altitude: 810 m), is 2337 mm. The 2300 mm estimate is, therefore, correct.

10. (pg 9167, L5) Please define INAG. Is it the Portuguese Water Resources Institute? I had to look this upon the web.

INAG was defined and now the inline reference includes The Portuguese Water Institute, INSTITUTO DA ÁGUA, I. P., (INAG, I.P.). Also, in section 2.4 the INAG dataset was described for a better understanding.

11. (pg 9167, L10-11) This is a key statement leading me to believe that the individual events during this month aren't actually that extreme and led to comment 1 above.

See response to #4.

12. (pg 9169, L17) Please define SNIRH and describe the dataset. There are also several references I was able to find online, although it does not appear that any of them were published. (<http://archive.rec.org/REC/Programs/Telematics/DETERMINE/WaterSession/JRibeira>)

Section 2.4 was rewritten to incorporate the definition of SNIRH and a brief description of the dataset was added.

13. (pg 9169, L18) Figure 1 does not include units on the elevation profile. Kilometers along the x-axis in the same panel would also be helpful.

Figure 1 was modified to include the proposed changes.

14. (pg 9169, L23-26) It might be helpful to linearly interpolate the model data to the observation location and compare that time series with the observations. This is another way of verifying just how much of the error is coming from the spatial issue or the timing issue. In areas of complex terrain this could make a huge difference.

The authors followed the referee's suggestion and interpolated the model grid points to

C5588

the station location. The anomalies from observations were computed and compared with the ones from the nearest grid-point series. The results showed no differences. A new paragraph was added to the "Methods" and to the "Results and discussion" sections.

15. (pg 9170, L9-14) The figure talked about in this paragraph does not exist and is the source of mismatched figure numbers. Figures 3-8 are all off by one

The mistake has been corrected. The numbering of the figures has been changed.

16. (pg 9170, L20-25) I feel like all of the statistical calculations show basically the same thing. Is it necessary to include all of them? The same goes for Figure 6, where panels a-e all show basically the same thing, the model run performance relative to each other. 15. (Section 3.2) Again, I think it is unnecessary to detail all of the various statistical methods. They appear to all be saying the same thing. Choosing one (or a couple) is sufficient to assess how the various runs compared to each other and the observations. A brief mention that you tried several other methods and they produced similar results would be sufficient.

The number of statistical indicators of model performance was reduced. The new version of the paper only shows the mean error, the mean square and the root mean square error, the correlation and the model skill. Since MSE and RMSE only provide information about the model accuracy, categorical measures have also been considered.

17. (pg 9171, L14) Do you mean to say overestimate instead of underestimate? 20. (pg 9176, L16) I think you mean overestimates, not underestimates.

In the new Section 2.5 the equations used were included for a better understanding of the results obtained.

18. (pg 9171, L16) Fig 3 should be Fig 4. Fig 4 should have units on the y-axis and perhaps a legend. Also, there is no mention of the colored plus symbols on the plot.

C5589

Either remove these symbols or address their meaning.

The Fig. 4 was changed and now is Fig. 5. The figures were modified to incorporate the proposed changes. The numbering of figures in the text, the units, colours and symbols were changed throughout. The captions were revised and the colored plus signs (outliers) were removed in order to make the figure more legible.

See responses to #5 and #15. 19. (pg 9172, L7) Fig 4 should be Fig 5. Please add units to the plot and the caption for Fig 5.

Please, see response to previous comment.

20. (pg 9172, L10) Fig 5 should be Fig 6. Please add units to the x- or y-axes. The plot is described as grey-scale in the caption, yet it is a color plot.

Please, see response to comment 12.

21. (Section 3.2) Again, I think it is unnecessary to detail all of the statistical methods. This section was rewritten to be in accordance with what is expected in a "Results and discussion" section. In the new version only the results are presented.

22. (pg 9173, L28) Do you mean climatology instead of persistence?

The mistake has been corrected. I do mean to say climatology.

23. (pg 9174, L1-2) The fact that the skills core indicates that some stations are better off with climatology instead of modelled precipitation is worrisome. What does this mean for the accuracy of representation of precipitation, etc., in the hydrologic models these runs are meant to be used for? This is never discussed.

Addressing the referee concerns about the model performance and model accuracy, the authors explored more deeply the formulation of the Mean Square Error Skill Score.

The skill score is defined by the ratio between MSE and the observed variance (MSE-clim). The MSE indicator is very sensitive to large errors. Large errors can be in-

C5590

terpreted as a loss of model accuracy. For the Mean Square Error Skill Score (SS), negative values indicate model performances worse than climatology but it does not necessarily imply that the model has no skill. A poor skill score does not reflect a bad model performance (Wilks, 2006; Jolliffe and Stephenson, 2003).

For the time and spatial scales that are studied in the present work and taking into consideration that precipitation is not a state variable but a highly parameterized one (both in the microphysics and cumulus parameterizations), probably the measure of model performance may not be the most suitable for the hourly precipitation.

(Wilks, 2006; Jolliffe and Stephenson, 2003) These references appear in the manuscript.

24. (pg 9174, L18) .the average magnitude of the error is the same: it is on average larger in areas of high slope and small in the lowlands."(pg 9174, L21-22) "magnitude error higher at the majority of stations and lower in the areas of rugged terrain (top right corner). These two statements appear to contradict each other. Especially since, in Fig 7c, it appears that the highest values on the plot are over the rugged terrain in the top right corner. The contour there says 2, which is the highest in the domain.

The manuscript was completely revised especially the "Results and Discussion" sections. Either by including a new analysis as by the remaking the results section. In this process the old Fig. 6 was removed and the mention contradictions are no longer applicable.

25. (pg 9174, Section 3.3) Figure 7d is never referenced. Is it necessary?

The results section has been modified and the mistake was corrected. The Fig. 7 mention was removed from the text.

Major Comments - Anonymous Referee #2 26. Title is incorrect. How the authors justify the term "extreme" in the title considering the relatively low return-values and the absence of particularly heavy daily events during December 2009. Also in P9164,

C5591

L6, what are extremely heavy rainfall periods? Indeed in P9164 L8-10 It is said that the climatological mean is exceeded in +89mm in average, but it makes a difference considering the range of regimes (190-1175mm).

Regarding the second part of the comment, the mixing of monthly anomalies and the total year amount of precipitation (in order to characterise the differences over the country) introduces some confusion for the reader.

The analysis from Table 1 had two objectives, one to show that the amount of precipitation during December of 2009 was higher than the median, for each of the stations presented, and on the other hand and combined with Table 2, show that the higher amount of precipitation accounted during this month was not due to shot/intense events but rather to moderate but prolonged in time precipitation events.

Also, see response to #1.

27. Despite the fact that the authors suggest that short-term extremes (30mins) are critical for surface runoff generation in Mediterranean scales, they do not address this issue in this paper (DO THEY?). Therefore, the authors should consider changing this sentence to emphasize the relevancy of their work or investigate short-term extremes.(P9165 L6-9).

Since this work was motivated for further hydrological applications, the sentence was introduced to give an overview of the aspects considered more relevant for the catchments in the region. Also, is intended to explain why the evaluation has been carried with hourly precipitation rather than the daily amounts.

The authors do not feel that the phrase was overweighted in the original text or the need for supplementary analysis. In line with the explanation given in major comment 4, in fact, there were no short term/intense events, and the authors have always assume that, but rather a great amount of rain during the month.

With the reorganization of the introduction the authors believe that the meaning of

C5592

phase has been clarified.

28. The introduction in general could be much clearer and more related to the topic itself (maybe more focused on modelling). I personally think that it could be reviewed and reorganised, setting up an appropriate scene for the study.

Following the referee comment the introduction was reorganized and it is now more focus on the modeling aspects of the simulation period. In addition to changes in the introduction we made corresponding adjustments to the abstract and discussion. Discussion was changed to incorporate the motivation and goal of the study.

See response to #6.

29. Several parts of the manuscript are presented carelessly. The manuscript contains several grammatical and orthographic errors, a large number of references are missing, figure's labels and references do not match and many of the figures lack information (e.g., units, scales). Some of those errors are detailed below, although not all of them.

The appropriate changes were made. In fact, all parts of the manuscript were reviewed. The figures and tables captions as well the legends were changed to include units, scales, meaning of the colored symbols and so on.

See response to #6.

30. Several statistics were computed to evaluate the model performance. The question that arises is whether all those statistics provide independent measure or additional information of the model performance. The respective results indeed suggest that they do not.

See response to #16.

31. Section 3.2 is in general confusing. It can be much improved. Several parts of this section are difficult to follow and the authors do not express themselves with clarity.

See response to #6.

C5593

32. The conclusions are a mere description of the results. They should not be only a summary of the previous sections. Different features that could be improved are described below. See response to #6.

Specifics Issues - Anonymous Referee #2

33. P9165 L15. Change northt by north. 3. P9165 L26-29. Parenthesis and full stops to be revised. 9. PG9167 L16. Replace "terrain following" by "terrain-following".

The misspelling was corrected.

34. P9165 L 22. What is the frequency of the model outputs? Temporal resolution of NWP models is much higher than 15-30 min. In particular, WRF requires a temporal resolution (in seconds) of about 6 times the spatial resolution (in km). That is, for a 1-km run, the time step should be 6 s. This is the timescale that the model is able to capture, although the modeller might have set up a lower frequency to write the outputs.

The referee addressed an important question that was overlooked. Thus, in the new version of the introduction a new paragraph was introduced with the aim of clarifying the differences among model resolution (model time step) and model integration times (recorded time of the models products).

35. P9165 L26-29. Parenthesis and full stops to be revised.

The phrase was rewriting.

36. P9166 L2-3. This is probably unnecessary. "due to battery failure as a result. . ."

It was re-written and removed. The idea was to point that data gaps may result from mundane explanations as lack of battery, due to lack of radiation because the skies were cloudier than average.

37. P9166 L12. "Sparsely" is an adverb.

It was replaced with "scarcity".

C5594

38. P9166 L10-18. I would rephrase the entire paragraph, but particularly the last sentence, which is too long and confusing. This is a key paragraph of the manuscript and should be neat and clear.

The sentence was rephrased. Now there are two paragraphs explaining the motivation, behind this study.

See response to #6.

39. P9166 L 24-25 Please, rephrase. ("Within the study. . .Águeda").

Changed to " In this region is located an important watershed, subject of recent studies, the Águeda river catchment (Figueiredo et al., 2009), well-known for its flooding risk to the old city centre of Águeda."

40. P9167 L 15. This reference is for version 3, not 2.2. (Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang and J. G. Powers, 2005: A Description of the Advanced Research WRF Version 2).

In the revision of the article and with the inclusion of new analyses the original model data were reviewed and it was found and that the model version used used in this study is in fact the 3.1.1. This has been corrected in the manuscript.

41. PG9167 L16. Replace "terrain following" by "terrain-following"

The change was made.

42. P9167. L 22. All except two references regarding WRF are missing in the bibliography. Despite not being wrong, the reference regarding the use of WRF in Portugal is a conference contribution, which is not accessible to readers (only the abstract). A quick search in any of the internet databases will return studies using WRF in Portugal or the Iberian Peninsula that could be cited and are probably easier to find by the reader. Also, some of the studies for the Iberian Peninsula (with WRF and other LAMs) could serve as comparison for the author's results and could also give clues on the source of

C5595

some of the errors.

The authors followed the recommendations of the referee, adding references to other works made for the same region. The references are now corrected.

43. P9168 L3-6 In the model setup, the authors should specify how the domains are nested (i.e., one-way or two-way).

The manuscript was completely redone. In the process the model setup was changed and the information was added. 44. P9168 L10-12. The location of the east border has not been selected according to this criteria. Indeed, according to figure 1, the east border of D03 crosses "Penhas Douradas", a point that is particularly complex in terms of regional modelling (see Fernández et al. 2007, JGR) The authors are aware of the problems that arise from choosing the east border crossing the "Penhas Douradas" location. However, in modeling one must account the balance between simulation time, horizontal resolution and the dimensions of the area of interest (number of grid-points). Running, an inner domain at 1-km of resolution, simulating a time period of the month, turns the system computationally expensive. So, the strategy was to expand the western border into the Atlantic Ocean to better capture the northwest predominant circulation, and the east border as far as possible from the Pousadas location (our interest rain-gauge location). In doing so, the eastern border was sacrificed.

45. P9168 L13-15. The eta levels can be omitted if they are the model default ones.

The manuscript was modified accordingly.

46. P9168 L15-22. The selection of the parameterization schemes is of paramount importance in a regional model. The author's should explain why this configuration was selected or provide references of previous studies if necessary. Also, the authors do not specify whether the convective scheme was switched off for the inner domain. At high spatial resolution (typically higher than 10km), the cumulus scheme should be switched off because the model becomes able to dynamically capture the convective

C5596

processes.

In this study the cumulus scheme was switched on for the inner domain. In the study performed by Luna et al. (2011), related with orographic precipitation, no differences were found between configurations with or without cumulus. Thus, the parameterisations used followed the work of Luna et al. (2011).

Luna, T., Rocha, A., Carvalho, A. C., Ferreira, J. A., and Sousa, J.: Modelling the extreme precipitation event over Madeira Island on 20 February 2010, *Nat. Hazards Earth Syst. Sci.*, 11, 2437-2452, doi:10.5194/nhess-11-2437-2011, 2011.

47. P9168 L36-28. This sentence is unclear ("They were. . .experiments"). I think that the explanation of the experimental design could be much more simple, although still containing the same information.

The experimental design was reformulated.

48. P9169 L14. As a general rule, humidity should not be nudged to avoid competition with convective schemes. However, it is not clear from the text whether the convective scheme is switched off (see comment 14) in the innermost domain. It is not clear either if the nudging is applied to all domains.

In the new version of the manuscript is now clear that the nudging was applied to all of the domains. The authors are not aware of the general rule that the referee mention on this comment. The nudging technique is a source/sink term that is included in the equations. It will definitely interfere with the balance of the overall variables. The works of Soares et al. (2012), Argüeso et al. (2011) and Fernández et al. (2007), apply the grid-nudging technique without referring the variables nudged, but the works of Lo et al. (2008), Xavier et al. (2008) and Chotamonsak et al. (2012) they refer that grid nudging was applied to humidity.

Chotamonsak, C., Eric P. Salathé Jr., Jiemjai Kreasuwan and Somporn Chantara, 2012: Evaluation of Precipitation Simulations over Thailand using a WRF Regional Cli-

C5597

mate Model. Chiang Mai J. Sci. 2012; 39(4) : 623-638. V. F. Xavier, A. Chandrasekar, H. Rahman, D. Niyogi, K. Alapaty, 2008: The effect of satellite and conventional meteorological data assimilation on the mesoscale modeling of monsoon depressions over India. Meteorol. Atmos. Phys. (2008) DOI: 10.1007/s00703-008-0314-7. Soares et al. (2012), Argüeso et al. (2011), Fernández et al. (2007) and Lo et al. (2008) – These references appear in the manuscript.

49. P9169 L19. I am not very familiar with the term "buddy". How are they determined?

The term is applied to neighbourhood stations and is most used in the meteorological field. Some of the references mention in this study used the term "buddy station" throughout. To avoid misinterpretations the "buddy" was changed to "nearby".

50. P9170 L9. Remove "the" before "December".

The entire section was rewrite and the sentence was replaced. See response to #6.

51. P9170 L13. The reader might think from the experiment design that the authors did not adopted a spin-up. It is specified later, but it is such an important feature of the simulations that it should be mentioned in the experiment design section. Did the authors consider longer spin-up periods?

As suggested by the referee, the experimental design section now includes the model spin-up time.

52. P9170 L10. Is this figure missing? I cannot see any time series in figure 2.

See response to #6.

53. P9170 L 16-18. I am not sure of having understood the procedure correctly. Does it mean that we are probably taking hours that are completely dry in the model and discarding wet hours just to maintain not only the same number of hours, but also the timing of those hours.

The procedure adopted consisted in series of simulated and observed matched pairs

C5598

with different lengths for each station, for observed hourly precipitation above 0.1 mm/h.

54. P9170 L27. I do not understand what the records evidence.

See response to #6.

55. P9171 L17. Why three times the IQR? Maybe reference or a justification.

The reference was included. The three times the IQR follows the suggestion provided by Wilks (2006) as a more robust measurement of the degree of dispersion among data than the use of maximum/minimum value.

56. P9171 L22-23. Do you mean among stations?

The sentence was revised. See response to #6.

57. P9172 L 6-8. Please rephrase these two sentences. "strength among data pairs"(?). Revise the tenses.

See response to #6.

58. P9172 L13-14. What is the sensitivity of the instruments? ME (and the other statistics) cannot have higher precision than the measurements themselves.

The original data is delivered with two significant digits for hourly precipitation records above the 1 mm/h and with one significant digit below the mentioned threshold. So, the error measurements do not have higher precision that the data itself.

59. P9172 L17-18. The sentence "Individually. . .control run" is unclear and should be rephrased.

See response to #6.

60. P9172 L18-20. I do not understand what the authors mean here.

See response to #6.

61. P9172 L23-26. I do not understand the evidence the authors try to show. These

C5599

sentences should be clearer.

See response to #6.

62. P9172 L27-28. Reorder the sentence "In Fig 5c. . . absolute error".

See response to #6.

63. P9173 L3-5. The authors included sentences like "Still. . . remaining stations" that are unclear to me. What do the authors want to express by "a few stations show an average magnitude error smaller than the reference run and that of the remaining stations"? If I have understood correctly, it is obvious that the model errors are smaller in some stations than in others. See response to #6.

64. P9173 L8-10. The fact that the authors use different statistical parameters to examine the errors do not necessarily characterize different aspect of the errors. They might be characterizing similar features and thus will agree.

See response to #16.

65. P9173 L13-15. If I have understood correctly (here and in many other places along the manuscript), the authors are asserting that some of the stations are more accurate than the aggregated statistic value. I do not consider this particularly interesting since it is in the very nature of an average.

The RMSE and MSE, and other non-linear indices, are calculated from the pooled sample and not averaged. Due to non-linearity the results are different.

66. P9173 L17. "aggregated", please revise tense.

See response to #6.

67. P9173 L21. Remove "only" since one of the simulations necessarily outperforms the others. Also change "outer perform" by "outperforms".

The changes were made.

C5600

68. P9173 The authors should define terms like "aggregated", "reference system", "reference run" and other, and use them in a systematic way to ease the reading and understanding of the manuscript. In different occasions, I was unable to know what the authors were referring to with these terms.

The Methods and Material section and the Results and Discussion section were completely redone. In the process the referee suggestions were taken into consideration. See response to #6.

69. P9174 and P9175. These sections are particularly difficult to follow because the respective figures are barely visible. The figures also lack information.

See response to #6.

70. P9175 L22. "The quality of the model. used as a measure of the quality of the model" (?).

See response to #6.

71. P9175 L24-25. There is no forecasting in the study. The authors should also tone down their statement "reliable and consistent" bearing in mind the results of their limited evaluation.

See response to #6.

72. P9175 L26. What do the authors mean by "the model system has resolution"?

The phrase was changed.

73. P9176 L 5-6. Why not choosing other measurements that take this into account?

See response to #16.

74. P9176 L7. The authors have not examined the suitability of the domain.

The referee is corrected. The phrase was changed.

C5601

75. P9176 L9-12. I do not see the relevance of this citation.

The quote was removed from the text.

76. P9176 14-15. Do the authors mean that some of the stations provide better results than the average of stations? As stated before, this is in the very nature of averaging.

See response to #65

77. P9176 L20-22. The authors have not tested any additional configuration with respect to the initialization. Also the reference cited actually proposed an alternative method to single initialization and concluded that frequent reinitializations might be beneficial. Also, please keep consistency in the format of the citations (see previous citation of Lo et al. 2008)

The paragraph in question was retired from the new version of the manuscript. However, Lo et al. (2008), when referring to frequent initializations with simulations with grid nudging, they did not find any advantages in using frequent initializations.

78. P9176. "However, proceeding with caution is necessary". Is it not a too broad statement to finish the manuscript?

The sentence was removed from the text.

79. All figures are barely visible and in many cases impossible to interpret because of their size and resolution. They also lack units and information. In figure 1, the location of the profile is not specified. Figure 4 includes markers that are not described and I am not able to see any boxes. Also rephrase the caption. It would be helpful to have a slope 1 line in figure 5 (as well as the units). The authors should make an effort to present the information of figure 6 in a way easy to interpret by readers.

See response to #6.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 9163, 2012.