Comment made by: Anonymous Referee #2

Received and published: 29 April 2020

Reply by authors is shown in blue, and starts with the symbol >>.

In their study, González-Rojí et al. investigate three different convective parameters obtained from two dynamically downscaled WRF model runs over the Iberian Peninsula. Over a 5-year period, the convective parameters from the WRF runs are quantitatively evaluated with sounding data and spatially investigated for different seasons. In addition, the spatial distribution and variability of the convective parameters is investigated and related to certain precipitation characteristics from the literature. The authors found that WRF runs with 4Dvar assimilation best reflect the convective situation.

>> We point the reviewer that we have used 3DVAR data assimilation.

Overall, the work is well structured and written with a good balance of text and figures. My main concern is that large parts of the paper are rather descriptive in the sense that mainly the figures are described and not interpreted. Reasons for the discrepancies found between the data sets are not given - although that would be most interesting and would increases the scientific value of the paper. In the current version, the benefit of the work for a larger community remains unclear. In the following you find a list of major and minor points as well as some suggestions for editing.

>> Thanks for your comments.

Major revision points:

1.) After reading the paper, more questions arise than answers or new scientific insights are given. This is because the paper mainly describes the figures, but does not provide explanations. Questions are: Why do the assimilation runs perform better compared to the simple WRF downscaling? Since the convective parameters considered depend on both temperature gradient and moisture, hat is better reproduced? On which levels/layers? Depending on the location (sounding station) and the season? Why are the differences between the models greater at some stations than at others (depending on the parameter)? What is the relation between CAPE and TT index?

>> The simulation including data assimilation produces more reliable results than the one without it. This conclusion is extracted from the paper after the analysis of the convective indices studied, after comparing the results from both WRF experiments against the ones obtained from Wyoming University (also against IGRA radiosondes as shown in one of the comments posted in the website).

>> The differences between WRF experiments are due to the effect of data assimilation in the the vertical profiles of temperature and mixing ratio. The effect of the assimilation is not restricted to the surface, and it is propagated towards the top of the atmosphere and the nearby grid points due to the optimization of the cost function (Barker et al., 2004, 2012). Additionally, as presented in previous studies (already cited in the manuscript in sections 2.1 and 2.2), the effect is also observed in the soil moisture and both surface temperature and moisture. As presented in González-Rojí et al.

(2018), data assimilation is important at 12 UTC for moisture, and at 00 and 12 UTC for temperature, and their effects are important in the southeastern IP and both Guadalquivir and Ebro basins (see their Figure 13). This pattern is consistent along the seasons, but its intensity varies seasonally (stronger during summer than in winter). As presented in González-Rojí et al. (2020), the soil moisture content is also different in both simulations as a result of the data assimilation (this variable is not assimilated, and data assimilation is the only difference in the configuration of the model).

>> The main objective of our paper is neither to find a relation between the studied convective indices over the IP nor their performance as predictors of heavy rainfall events. We only want to evaluate how well the values of each index are simulated by comparing the results from two different configurations of the model to observational data, and to study the differences in the seasonal patterns due to the use of a data assimilation step in the numerical downscaling phase. There are not many studies analyzing this currently.

>> ----------

>> González-Rojí, S. J., Sáenz, J., Ibarra-Berastegi, G., & Díaz de Argandoña, J. (2018). Moisture balance over the Iberian Peninsula according to a regional climate model: The impact of 3DVAR data assimilation. *Journal of Geophysical Research: Atmospheres*, *123*(2), 708-729.

>> González-Rojí, S. J., Sáenz, J., Díaz de Argandoña, J., & Ibarra-Berastegi, G. (2020). Moisture Recycling over the Iberian Peninsula: The Impact of 3DVAR Data Assimilation. *Atmosphere*, *11*(1), 19.

>> Barker, D. M., Huang, W., Guo, Y. R., Bourgeois, A. J., & Xiao, Q. N. (2004). A three-dimensional variational data assimilation system for MM5: Implementation and initial results. *Monthly Weather Review*, *132*(4), 897-914.

>> Barker, D., Huang, X. Y., Liu, Z., Auligné, T., Zhang, X., Rugg, S., ... & Demirtas, M. (2012). The weather research and forecasting model's community variational/ensemble data assimilation system: WRFDA. *Bulletin of the American Meteorological Society*, *93*(6), 831-843.

2.) The main conclusion of the paper is that the assimilation run performs better compared to the run without assimilation. But is this not to be expected if soundings are assimilated for which the comparison is made afterwards? What would be the result if you left out some of the soundings for the assimilation and made the comparison for these locations?

>> The paper supports the idea that the experiment including data assimilation performs better than the one without, similar conclusion to what we have observed for other variables in previous studies by the authors. However, in this case, the main conclusion of the paper is that important differences arise in those patterns only due to data assimilation. The impact of data assimilation is not limited to the grid cells close to the location of the soundings. As shown in the Figures of our paper, the changes extend over large areas of the Iberian Peninsula despite the limited coverage by soundings.

>> It is true that the comparison against assimilated soundings can be biased, but we can not discard observations when preparing the simulations without performing a damage to the study that we want to perform. On the other side, as we mentioned before, we are analyzing derived variables not directly assimilated on a regional domain covering places with no observation at all. We are mainly comparing the values of different convective indices after different calculation methods (as the

method followed by Wyoming and our method included in the package aiRthermo). Additionally, as an extra way of validating our results, we always compared the values obtained in the patterns over the entire IP with previous studies focusing in the region (or at least covering it even if it is with low resolution data).

3.) Are you sure that ERA-Interim did not originally assimilate the eight soundings you considered? It does not make sense to assimilate any data set twice.

>> We did not check every cycle of six hours all the data assimilated by ERA-Interim, as we think it is pointless. We actually assume that some of these radiosondes have very likely already been assimilated in ERA Interim reanalysis. However, that is not a problem for our simulations with WRF as we only used the data from ERA-Interim as boundary conditions for our regional model after the initial run. Since the run which used ERA Interim for initial conditions (January 1st, 2009) corresponded to one year before the period that we started analyzing the output (January 1st, 2010), we can be sure that the interior of the domain is reflecting the variability corresponding to the regional climate model.

>> Moreover, the effect of assimilating one station in ERA-Interim, which has a resolution of around 80 km, cannot be comparable to the effect of assimilating a station in a domain with 15 km. Besides that, the original objective of our paper was to compare the quality of WRF simulations and ERA Interim is only used to provide initial and boundary conditions to WRF.

4.) Either there is a general misunderstanding of convection triggering or the formulations are clumsy. Convective instability and sufficient moisture at lower levels are necessary but not sufficient conditions for the development of convective storm. Convection initiation requires additionally a lifting mechanisms that either reduces CIN or lift a parcel to the level of free convection (LFC). High CAPE/TT values neither trigger convection nor can they directly be related to precipitation as written several times throughout the manuscript.

>> To some extent, we agree with the reviewer. Convective instability and moisture in low levels of the atmosphere are ingredients necessary to trigger convective storms, and consequently, convective precipitation. However, the final ingredient, which is the lifting, is provided by the instability, forced by orography, the convergence of horizontal moisture fluxes or the breezes in coastal regions. All this information is included already in the second paragraph of the introduction of our paper, so we agree with the reviewer on that.

>> In order to avoid misleading ideas by the readers, we have carefully rewritten all the sentences highlighted by the reviewer in the new version of the manuscript.

5.) CIN works only in conjunction with CAPE. In case of zero CAPE, CIN doesn't matter for convective initiation or development. Analyses of the mean values or the spatial distribution of CIN are useful only when considering days with a certain amount of CAPE (or instability in general).

>> We agree to some extent with the reviewer on that. However, the objective of this paper is not to evaluate CAPE and CIN only for extreme events as tools to predict extreme convective rainfall. The objective of this paper is to evaluate the ability of WRF simulations (including or not the 3DVAR data assimilation step) to produce reliables values of TT, CAPE and CIN over the Iberian Peninsula, irrespective of whether they produce or not rainfall events.

>> As stated already at the end of the Introduction, "the main objective of this paper is to evaluate the performance of two simulations created by using the WRF model at reproducing the atmospheric conditions that can trigger convective precipitation over the IP. To do so, the comparison of pseudo-soundings extracted from the model against real observations will be carried out." At the very end, what we are doing in the paper is to evaluate the Probability Density Functions (PDFs) of the three instability indices obtained in each experiment against the reference values measured by the University of Wyoming (also IGRA in the future version), but not only during extreme events.

>> This clarification was added to the new version of the manuscript.

6.) Using only the nearest grid point to a sounding station neglects the horizontal drift of the radiosoundings. A better choice would be to consider the average value of several grid points.

>> That is true to some extent. We agree that considering the nearest grid point for the comparison against a sounding is not always the best option. However, this depends on the spatial resolution of the domain of the simulations. Averaging several points can be a good idea when convection-permitting scales are used (below 5-3km), but not when the spatial resolution of the experiments is 15 km (as in our case). If we consider the average of the nearest grid points, we would be taking into account an area of 2025 km2 (45km x 45km), and that is too much for a comparison against station data.

>> Additionally, according to recent studies (Xu et al., 2015), most of the vertical levels up to 6 km are already measured for a drifting distance of 7.5 km, independently of a clear or cloudy day (see their Figure 6). As also mentioned by the reviewer, both convective instability and sufficient moisture at lower levels are necessary for developing a convective storm, and these lower levels are already measured below 6km. Thus, taking into account our spatial resolution, we stand by our decision to use the nearest point to the station for comparison against station data.

>> ----------

>> Xu, G., Xi, B., Zhang, W., Cui, C., Dong, X., Liu, Y., and Yan, G. (2015), Comparison of atmospheric profiles between microwave radiometer retrievals and radiosonde soundings, *J. Geophys. Res. Atmos.*, 120, 10,313– 10,323, doi:10.1002/2015JD023438.

7.) No reference is made on the original ERA-Interim fields. Thus it is not possible to assess the added value of the downscaled model runs and the need for higher resolutions of the data.

>> The information about the original ERA-Interim fields was also asked by the comment published by a reader. As stated in his reply available online, we used 20 pressure levels downloaded from the MARS repository to feed the WRF model, which are: 5, 10, 20, 30, 50, 70, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 925, 950 and 1000 hPa. Our set-up, as stated in the manuscript, uses 51 vertical levels, so there is a relevant increase in the number of vertical levels compared to the data from ERA-Interim. Additionally, the spatial resolution of ERA-Interim is around 80 km, and our domain has 15 km resolution. Thus, we also improve the spatial resolution of the data.

>> Taking into account this information already stated in the manuscript, we sincerely consider our simulations provide extra information to the one present in the Reanalysis. Additionally, these two

experiments have been already validated against observational datasets (both for stations and grids) in previous studies by the authors, and in some cases, particularly for the experiment including data assimilation, they are able to outperform the driving reanalysis ERA-Interim. All these studies are cited at the end of section 2.1. We have not performed any quantitative analysis of the added value of these simulations since, as we have already stated before, we do not compare the performance of the WRF runs with the original data (see Figures 2 to 8 of the original manuscript, for instance). We are interested in comparing the performance of a run using 3DVAR with a different one which does not use it.

8.) The last section "Conclusions" is only a summary without any (general) conclusions. Tell us what other scientists may learn from your study.

>> We do not agree with the reviewer. It includes all the important information extracted from the analysis performed, and it includes details about the comparison of both experiments regarding the indices TT, CAPE and CIN, not only in the location of the radiosondes but also for the entire IP.

9.) A thorough language check is necessary (e.g., "...observations **in** the stations..." or "obtained **in** stations" or similar formulations used throughout the manuscript are incorrect/weird).

>> A detailed revision and edition of the language has been carried out in the new version of the manuscript.

Minor revision points:

1. Explain why you have selected CAPE, CIN (note my comment above), and TT and not others, in particular indices that either estimate potential or conditional instability or dynamical properties (deep layer shear, storm-relative helicity; or an index combining thermodynamical and dynamical properties). Is there any cross-correlation between those parameters (CAPE vs. TT)? Also explain why you have only considered a 5-year period, which is far from being representative for the general climate.

>> As stated in the manuscript, we considered some of the most commonly used convective indices, which can give information about the regions where more unstable conditions are met over the IP. This is not something weird or new, and that is why several studies focus only on some of these indices, or only even in one of them. Some examples of these papers, and particularly focusing in the IP, can be found in the Introduction of the manuscript.

>> About the length of our simulations, in any case we say in the manuscript that we want to show a climatology of these indices, as we also agree that it would be impossible only with 5 years. As presented in the paper, we only want to evaluate the differences triggered by the use of data assimilation in those patterns for a limited period of time. Since the same period of time is used for both simulations, and since the same model, parameterizations and boundary conditions are used for both runs, the differences identified must be clearly assigned to the use of 3DVAR data assimilation. The period of time is shorter than the estimated 30 years needed to robustly resolve the climatology, but it is long enough (five years is not a week) to draw robust conclusions across the behaviour in different seasons of the year, for instance.

2. It's very difficult to compare the different sub-figures due to different axis ranges. I suggest to using the same scaling within one figure.

>> We agree on this comment with the reviewer because having the same scaling in the figures is easier to interpret the results, particularly for the intercomparison of results. However, that is not possible in our case because the values show a really large range. Here are some examples concerning each of the figures included in the manuscript:

>> 1) Figures 3 and 4 (Taylor diagrams for CAPE and CIN): A Coruna presents standard deviations of around 25 J/kg for CAPE, but Barcelona presents values around 250 J/kg. If we set the axis to the maximum, the results from many stations will not be recognizable. The same happens with CIN, as Gibraltar and Murcia present values around 100 J/kg, and A Coruna around 18 J/kg.

>> 2) Figure 5 (Box and Whiskers for TT) already has the same axis range.

>> 3) Figures 6 and 7 (Box & Whiskers for CAPE and CIN): most of the values for CAPE in winter are below 75 J/kg (and some stations show values around 0 J/kg), but in summer some stations reach the 750 J/kg. Same happens to CIN, in which all the stations obtained values below 20 J/kg in winter, and below 400 J/kg in summer.

>> 4) Figures 9 and 10 (patterns for CAPE and CIN): Same thing as before happens. For CAPE, the values in winter are below 50 J/kg, but in summer are below 600 J/kg. For CIN, the values in winter are always below 10 J/kg, and in summer below 250 J/kg. However, in these Figures, the same axis range has been selected for each season, independently of the time of the day in order to clarify the results.

>> In order to set the same range of values for the TT index, which results do not vary as much as for CAPE or CIN, Figures 2 and 8 will be created again. However, the other figures remained the same in the new version of the manuscript as we truly believe that setting the same axis for all the plots in each Figure will complicate the visualization and interpretation of the results.

3. When describing the general convective situation over the IP / over Europe, you should consider also more recent literature.

>> Some of the most recent papers focusing ONLY the IP were presented in the introduction, and that is why even the ones focusing in future scenarios were commented in there.

4. Why have you created your virtual WRF soundings only from one grid point? As correctly stated in the text, the soundings may drift over some distance during the ascent. Using an array of 3 x 3 grid points or so would have been a better choice. Please add a comment on that.

>> As stated in the mayor comment number 6 of the reviewer, that would be necessary if convection-permitting scales were used in the simulation. However, the spatial resolution that we used is 15 km resolution, and most of the levels measured by a balloon are already measured when the drifting distance is below 7.5 km (the height of the balloon is around 6 km) independently of the conditions in the sky. This distance is less than the distance covered by our grids, and increasing the grid to the nearest points (15km x3 grids) will not be a good option to evaluate the performance of the model.

5. L1 (see major point above): Instability does not trigger convection.

>> We have edited that sentence as suggested by the reviewer.

6. L2 (also L29-30): CAPE/CIN are measures of the energy and not instability indices.

>> True. CAPE and CIN represent the Convective Available Potential Energy and the Convective Inhibition energies in a column of the atmosphere, but they are related to atmospheric instability. See, for instance, Tsonis (page 155), in which CAPE and CIN are discussed as problems in Chapter 8, entitled "Vertical stability of the atmosphere". In Djuric (1994), Chapter 5 is entitled "Analysis of vertical soundings" and section 5-6 is entitled "Instability Indices" and Section 5-8 is entitled "Integrated indicators if instability". These sections describe the indices we present in our paper. Bohren and Albrecht (1998) write (we quote): "A sounding (not just a layer) is often said to be conditionally unstable or in the conditional state if a parcel from any level of the sounding has positive CAPE" in page 317.

>> ----------

>> A. Tsonis, (2007), "An Introduction to Atmospheric Thermodynamics", 2nd Ed., Cambridge University Press.

>> D. Djuric (1994), "Weather Analysis", Prentice Hall

>> C. F. Bohren and B. A. Albrecht (1998) "Atmospheric Thermodynamics", Oxford University Press.

7. Shorten the abstract and focus on the essentials.

>> The abstract was shortened according to the suggestion by the reviewer.

8. L14: "the ingredients for the development of convective precipitation": As alluded to previous, you investigated only the convective environment, thus only one part of the ingredients.

>> The sentence was edited to incorporate what the reviewer said.

9. L22-23: Do you mean warm fronts? Note that cold fronts especially during summer frequently trigger convective storms by cross-circulations. Thus, classifying precipitation into frontal and convective does not make sense. Convective precip is not triggered by convective instability (see major point 4).

>> As stated already in those lines, we follow the definition by the WRF model of considering precipitation separated in two components: large-scale and convective precipitation. We state explicitly the frontal systems only as an example of that kind of precipitation, but that does not mean that we only consider that in that category. In order to clarify that, we have rewritten those lines.

10. L24: "The latter is usually associated with extreme events due to their intensity and short duration". Convection is per se not extreme. And you may add here "high intensity". But the short duration is not the reason why convection may become extreme (or rather related precip and wind).

>> We agree with the reviewer that convection is per se not extreme, and that is why we have already stated that usually convective precipitation can end up in a extreme event.

11. L24-26: The limited skill of NWP models to reliably simulate convective precipitation is not because of their low resolution (note that several European weather services run their models

already at 1 km resolution), but partly caused by forecast errors on the synoptic scale, which drive the predictability of convection initiation, and various sources of uncertainty on small scales such as limitations in the assimilated observables or microphysical schemes. There exist a bunch of literature on that.

>> We have added these other limitations to the text as suggested by the reviewer.

12. L37: It is impossible to estimate the life cycle or the intensity of convective storms from thermodynamic quantities solely. For organized convective storms, which represent the most intensive storms, you require sufficient vertical wind shear – speed shear (crosswise vorticity) for multicells, and directional shear (streamwise vorticity) for supercells.

>> That line highlights the ability of both CAPE and CIN to provide some information about the potential development and intensity of convective precipitation. As stated later in the same paragraph, we stated that another extra ingredient is needed to trigger it, such as the lifting.

13. L41-42: You state that "a (high; include) spatial and temporal resolution is important" for resolving vertical lifting, and thus regional simulations are needed. But in your study, you investigate only the convective environment and not the mechanisms relevant for convective initiation. So I do not see why you need higher resolved met. fields.

>> That line is not related only to the fact that high resolution is needed for resolving vertical lifting (that is obvious). As stated there already, high resolution data is needed to carry out similar studies to those presented in the paragraph, which evaluate different convective indices (as our study).

>> Additionally, the spatial resolution is important because in order to calculate variables as CAPE and CIN, these are more reliable when the resolution is finer. Particularly when you want to validate those results against radiosonde data or in areas of complex topography. As already mentioned, it does not make sense to calculate these variables with the mean values over huge areas or several grid-points. As shown in our paper (Figures 8 to 10), the spatial variability of TT, CAPE and CIN strongly resemble the features of terrain.

14. L43-44: The reason why convection peaks in the afternoon is related to solar irradiation. This is a fact and not "suggested by previous studies".

>> We agree on that, and that is what is expected. However, as stated later in the same paragraph, we show that some regions show those peaks in the morning. Thus, as it is not true everywhere, we used "suggested" to introduce that feature. We will change it to "backed-up" in the new version of the manuscript.

15. L44-45: Van Delden used only Synoptic stations with a 6-hourly resolution for their statistics. He found that "most thunderstorms occur at 18 and 24 UTC". 18 means the period from 12 to 18. Thus thunderstorms are most frequent between 12 and 18 UTC! But: It would be better to cite more recent studies based on lightning detections such as, for example, Piper and Kunz, 2017 (Nat. Haz. Earth Syst. Sci.; Fig. 4), Enno et al., 2020 (Atmos. Res.; Fig. 9), or also Lopez et al., (2001), the latter already cited. Not also that Corsica is not the only exception showing a different diurnal convection cycle (e.g., Fig. 4 in Piper and Kunz, 2017).

>> Some of the papers highlighted by the reviewer were added to the new manuscript, and the lines addressing the foundings by van Delden were adapted.

16. L50: Kaltenböck et al. (2009) investigate the relation between convective environment, lightning data and severe storms reports only for Europe, but not for the USA. So replace the citation or delete this statement.

>> That is true, Kaltenböck et al. (2009) only focuses on Europe. However, when he focuses on CAPE (in section 3.4 of his paper), he does the next statement: "Reasons could be the small sample or an underreporting of F2 and F3 events, the synchronous occurrence of different severe events (e.g. tornado accompanied by hail) and standard values of CAPE and SRH, which seem to be lower for Europe than in the US." And that is why that paper is cited in that line.

17. L51-56: The discussion of the convective environment should consider more recent publications based on lightning or high-resolution climate models (e.g., Mohr et al., 2015 (GRL); Sanchez et al., 2017 (Atmos. Res.); Rädler et al., 2018 (JAMC); Enno et al., 2020 (Atmos. Res.)).

>> As already stated in the reply to comment 15, some of the papers highlighted by the reviewer were added to the new manuscript.

18. L57: Explain how the seasonality of precipitation is determined by topography?

>> The reviewer is right. Precipitation is not determined by topography. In the new version of the manuscript this line was modified in order to highlight the fact that the seasonal precipitation patterns are affected by several factors, including: Different sources of moisture due to seasonal variations of the global atmospheric circulation and contrasting climatic regions (influenced by the strong topography of the Iberian Peninsula).

19. L64-65: These are not very high values for CAPE. On single days, they can be much higher in the interior of IP (note that according to Fig. 6. monthly mean has a maximum at 1250 J kg-1, which implies that at single days much higher values than 1000 J kg-1 are reached).

>> In that paragraph we are not evaluating if the values are high or not. We are just presenting to the reader the mean values of CAPE obtained over the Iberian Peninsula for a season, and the differences between the north and the south. We have rewritten that sentence in order to clarify it.

20. L72-73: These are very low values. Other studies (e.g., Kunz, 2007, NHESS; Pucik et al., 2015 (MWR), Taszarek et al., 2017 (MWR)) found much higher CAPE values (also for different version of the mixed-layer CAPE). This should at least be mentioned.

>> As we said in the previous comment, we only present the mean values obtained by previous studies in those lines. We have also included in the new version some of the other values included in the papers suggested by the reviewer.

21. Paragraph 57-75: Separate between precipitation and convective environment (CAPE).

>> Those lines were separated in two different paragraphs in the new version of the manuscript.

22. L76-80: I do not see how climate change is related to this work. I propose to delete this paragraph.

>> We wanted to show that indices like CAPE have been also investigated under future climate change scenarios. That is why we added that paragraph. We think that it is important to show that, so we have reduce it and merge it with previous paragraph instead of completely deleting it as suggested by the reviewer.

23. L81-86: Please better explain the objectives of the work. Evaluation is not an aim, but a method. Why is the evaluation of the convective field of interest?

>> We do not agree on that with the reviewer. We want to evaluate the performance of both WRF downscaling experiments at simulating some of the commonly used convective indices, and observe the differences that arise due to the use of data assimilation in one of these experiments. Taking into account all the information given in the introduction, it is clear that this topic is quite important to evaluate the regions more prone to develop unstable thermodynamic conditions that can end up in convective precipitation. Additionally, the importance of this topic is backed-up by all the papers that can be found in the literature, and particularly in our paper, by all the mentioned studies focusing over the Iberian Peninsula. Additionally, we want to stress that most of the WRF runs being currently run do not use the 3DVAR assimilation step, and we think that showing that it allows a better estimation of CAPE or CIN is an important contribution to the literature.

24. L110: Give some more details on the levels: spacing, highest level, which ones are used to compute CAPE/CIN.

>> This is something already asked by one of the readers of the paper, who posted a comment during the open discussion. As we stated in our reply to him, 51 vertical levels are available in our WRF experiments, and they go up to 20 hPa. In WRF, these vertical levels are in η coordinates, so they follow the terrain of the domain. Thus, the spacing between them is not constant. Explicitly, these are the values:

>> 0.9965, 0.988, 0.9765, 0.962, 0.944, 0.9215, 0.8945, 0.8649009, 0.8347028, 0.8045048, 0.7743067, 0.7316024, 0.6780097, 0.6275734, 0.5801385, 0.5355568, 0.4936861, 0.4543901, 0.4175383, 0.3830059, 0.350673, 0.3204254, 0.2921534, 0.2657521, 0.2411216, 0.218166, 0.1967937, 0.1769174, 0.1584536, 0.141405, 0.1258691, 0.1118248, 0.09912901, 0.0876521, 0.07727711, 0.06789823, 0.05941983, 0.05175545, 0.04482694, 0.03856365, 0.0329017, 0.02778335, 0.02315643, 0.01897375, 0.01519264, 0.01177457, 0.00868467, 0.005891433, 0.003366379, 0.001083758.

>> For the calculation of the indices in both WRF simulations, all the available pressure levels were used. As we replied to the reader, all this information will be clarified in the corresponding sections of the paper: 2.1 and 2.3.1.

25. P5, 1st paragraph: This part is a bit out of context in the section “Data and Methods”. Consider to move it to the introduction.

>> We do not agree with the reviewer on that. All the information given in this paragraph is related to the previous analyses and validations of the simulations that are going to be used in the study,

that is, experiments N and D in the paper. That is why this paragraph is presented here after the short introduction of both experiments.

>> We believe that if we move this paragraph to the Introduction, it will be completely out of context as the Introduction is mainly focusing on previous studies about the topic of the paper, that is, instability indices.

26. Section 2.2: Why haven't you considered IGRA sounding data?

>> This is something also suggested by one of the readers of the paper, who posted a comment during the open discussion. As presented in our reply to him, we used the data from the University of Wyoming because we used already the data in González-Rojí et al. 2018 for the validation of the precipitable water, and in that case, none of the values were taken as erroneous. Additionally, we wanted to keep a consistency between all the studies carried out with both WRF simulations, and that is the reason why the same radiosondes were used for the calculation of the instability indices in this paper.

>> In the reply to the reader, we also validated the indices calculated by Wyoming against the ones calculated by IGRA. As CAPE and CIN are very sensitive to methodological factors such as the computation of the initial parcel or the vertical spacing in pressure levels, we estimated them from the values of temperature and relative humidity at pressure levels in IGRA soundings using the same methodology that we have used in our paper (see Section 2.3.1). The comparison shows that the results are not sensitive to the selection of the dataset (see enclosed Taylor diagrams). We expected these results, since the use of homogenous data is particularly important for long-term trends, and we are simply analyzing five years of data.

>> Finally, as stated in the reply to the reader's comment, our results are robust to the selection of the observational dataset. However, as we feel that comment leads to a better paper, the figures included in it will be incorporated to the final version of the paper.

>> ---

>> González-Rojí, S. J., Sáenz, J., Ibarra-Berastegi, G., & Díaz de Argandoña, J. (2018). Moisture balance over the Iberian Peninsula according to a regional climate model: The impact of 3DVAR data assimilation. *Journal of Geophysical Research: Atmospheres*, *123*(2), 708-729.

27. L135: For readers outside of Europe it would be helpful to include here also local times (approximately).

>> The local times were added to that line.

28. L146: what is meant by "...the analysis increments are stronger at 12 UTC..."? And by "Strong increments are observed during summer..." in L148? Also the relation to the cold-bias in L149 is unclear.

>> The effect of the data assimilation is measured by the analysis increments (analysis minus background) at the analysis times (00, 06, 12 and 18 UTC). We analysed these quantities In González-Rojí et al. (2018), and we showed that the effect of the data assimilation was more intense at 12 UTC compared to the other times, and particularly for summer. The spatial analysis of these

values highlighted that the effect of data assimilation is not homogeneous over the Iberian Peninsula, and it concentrates mainly in the southeastern IP and both Guadalquivir and Ebro basins. This is related to the well-known cold bias observed in the IP in summer in WRF simulations, as the data assimilation is able to make it much smaller.

>> All these lines were edited to clearly state what is explained in the previous lines.

>> ----------

>> González-Rojí, S. J., Sáenz, J., Ibarra-Berastegi, G., & Díaz de Argandoña, J. (2018). Moisture balance over the Iberian Peninsula according to a regional climate model: The impact of 3DVAR data assimilation. *Journal of Geophysical Research: Atmospheres*, *123*(2), 708-729.

29. L150: "...the effect of the assimilation is not restricted only to the station location". This is a very crucial point. Unfortunately, you did not show that. see major point 2

>> That line is not related at all with the conclusions of this paper. As stated there, we are highlighting the fact that the data assimilation not only affects the nearest point to the observation, but it also is able to affect the nearest points as its effect is propagated zonally, meridionally and vertically. That is something expected since the main goal of the background error covariance matrices is to define how the effect of the data assimilation is propagated to the near cells.

>> Regarding the major point 2 from the reviewer, as stated already there, it is true that the comparison against assimilated soundings can be biased, but we are analyzing derived variables not directly assimilated on a regional domain covering places with no observation at all. Additionally, comparing station data against the mean of the 3x3 grids over the Iberian Peninsula would be ideal for convection permitting scales, but not for our 15 km resolucion domain.

30. Sect. 2.3.1: Please explain briefly how you compute the lifting curve from the surface/mixed level to the LCL to the LFC and to the LNB (including quantification of e).

>> All these information is already included in Sáenz et al. (2019) and also in the manual of the R package aiRthermo associated to that publication (available in CRAN). As stated there, "To compute CAPE and convective inhibition (CIN), the vertical integrals are computed in pressure levels by adding the energy corresponding to discrete slabs defined by linear or logarithmic vertical profiles, which are defined by the soundings. The integrals for each of the slabs enclosed by linear profiles are computed analytically, and the energy corresponding to each slab is accumulated, producing the final value of CAPE or CIN. The integrals are always calculated using the virtual temperature (Doswell and Rasmussen, 1994).

>> There are different methods of accurately determining the lifting condensation level (LCL) or the equivalent potential temperature of an air parcel in aiRthermo. In the first case, the package calculates these variables by computing their vertical evolutions and numerically solving the ordinary differential equation representing their ascent from the initial conditions given by their temperature, pressure, and mixing ratio. For compatibility, functions that allow these variables from well known alternative equations to be computed, such as the approximate method presented by Bolton (1980) to compute LCL, are also provided."

>> This information was already summarized in lines 162-165 of the previous version. These lines have been rewritten to make it much clearer in the new version.

>> ----------

>> Doswell III, C. A., & Rasmussen, E. N. (1994). The effect of neglecting the virtual temperature correction on CAPE calculations. *Weather and forecasting*, *9*(4), 625-629.

>> Bolton, D. (1980). The computation of equivalent potential temperature. *Monthly weather review*, *108*(7), 1046-1053.

31. L158: Do you have any reference for the statement that soundings "take many minutes to measure the profile of the atmosphere"? The multiplicity of soundings I performed in the past took ~ ½ hour to reach the LNB.

>> We think that 30 min could be defined as "many minutes". Anyway, that line has been edited to clearly state that the measurements made by the soundings are not instantaneous.

32. L169: What is "an isobaric precooling" and why was it applied?

>> As stated in that paragraph, in order to follow a similar methodology to that used by the University of Wyoming, the averaged values from the lowest 500 m were considered and an isobaric precooling was applied to the initial parcel state.

>> As shown by references provided in the paper, the computation of CAPE and CIN is very sensitive to the characteristics of the initial parcel that is lifted. We decided to follow the procedure of averaging the lower levels recommended by Craven et al. (2002). However, During the development of aiRthermo, we realized that the use of a parcel averaged for the low layers of the atmosphere could very often lead to underestimations of CIN. The reason is that it can happen that after averaging the lowest levels of the sounding to compute the initial parcel, it is still too hot compared to the ambient conditions. In that case, CIN will never be computed, since the initial parcel is already (artificially) buoyant. Thus, a cooling must be applied to the parcel if it is warmer than the environment. In aiRthermo, two options are available: adiabatic precooling (adiabatic ascent until the lifting parcel crosses the sounding) or isobaric precooling (the parcel is cooled along an isobar until it crosses the sounding so that it is not buoyant)). In our study, the isobaric precooling was chosen.

>> ----------

>> Craven, J. P., Jewell, R. E., and Brooks, H. E. (2002). Comparison between Observed Convective Cloud-Base Heights and Lifting Condensation Level for Two Different Lifted Parcels, Weather and Forecasting, 17, 885–890.

>>Sáenz, J., González-Ro jí, S.J., Carreno-Madinabeitia, S., & Ibarra-Berastegi, G. (2018). Manual for the R package aiRthermo. https://cran.r-project.org/web/packages/aiRthermo/aiRthermo.pdf

>> Siedlecki, M. (2009): Selected instability indices in Europe, Theoretical and Applied Climatology, 96, 85–94.

33. L172: As TT relies on temperature differences, the unit (°C, K) does not matter.

>> We agree on that with the reviewer. We just wanted to state the definition provided by Miller in 1975. We have edited that line in the new version.

34. L173-174: You should mention that other authors found other values (e.g., Huntrieser et al., 1996: 46 K; Haklander and Van Delden, 2003: 46.7 K; Kunz, 2007: 48.1 K).

>> These references have been added to the text.

35. L174-176: "...not highly dependent on the initial conditions..." correct (but even absolutely independent), but why differ the values you compute by your own from those provided by Wyoming – and only at Murcia (strongest), Santander, Zaragoza, Barcelona? Prevailing inversion layers as stated in L176 cannot be the reason as TT is based on main pressure levels which are always provided by Wyoming. Considering the initial values, you stated that you used the same mixing over the lowest 500 m, thus your method must be identical to that of Wyoming. How have you determined the LCL/LFC?

>> The differences are due to the different methodologies in the calculation of the TT index, mainly due to the Dew Point temperature. In the case of the University of Wyoming, the value is taken directly from the measurement. However, in our case, we do not use that measured value and we calculate it with the Pressure, Temperature and Mixing ratio at 850 hPa.

>> Figure 1 shows the scatterplots of the monthly means of the observed TT (computed from the values obtained directly from Wyoming University), and the ones computed with aiRthermo. In all of the cases, the R2 is above 0.99 for all the stations with the exception of Murcia (0.96). The worst values are obtained in the stations highlighted by the reviewer: Murcia (strongest), Santander, Zaragoza, Barcelona. Particularly in those stations, the values obtained by the University of Wyoming are larger than the ones obtained by aiRthermo from measured pressure, temperature and mixing ratio. The differences between the values are small, but they affect the correlation.

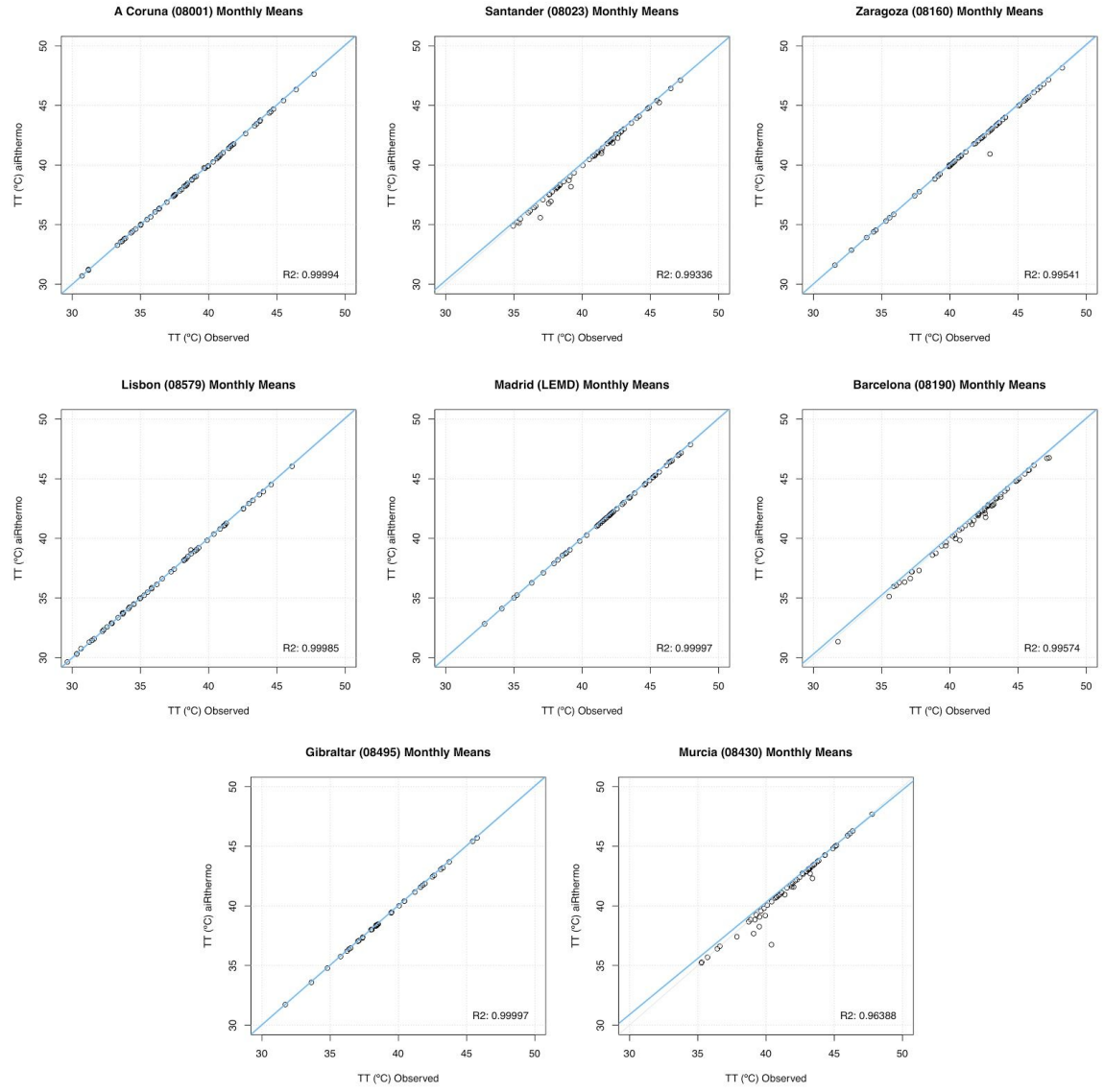

Figure 1: Scatterplots for the monthly values of the observed values of TT (directly taken from the University of Wyoming) against the monthly values of TT calculated with aiRthermo. The R-squared is presented in the bottom right corner of each plot.

>> We do not believe that this Figure should be added to the manuscript, but some details about these results will be given in the text after the Taylor diagrams for TT.

36. Sect. 2.3.2 / Results: The statistical distribution of CAPE is highly skewed. The product moment correlation coefficient according to Pearson, however, require a normal distribution. A better choice would be the rank correlation coefficient according to Spearman.

>> Tailor diagrams can only be created by means of Pearson's correlation. In order to show visually which experiment is best compared to the reference data, we decided to sacrifice the use Spearman correlation. The reason for this is that Taylor diagrams are built considering the mathematical relationship between the correlation coefficient, the RMS error and the variance of the series.

>> However, we show here the comparison of the bootstrap results for both correlation types. Figure 2 shows the results for CAPE and Figure 3 the results for CIN. If we focus only on Figure 2, we can see that the values are similar in most of the stations and only in A Coruna and Santander there is a strong worsening of the values. In any case, the same structure is observed: aiRthermo is the closest one to the values obtained by IGRA, followed by Wyoming, WRF D (experiment including data assimilation) and finally by WRF N (without data assimilation).

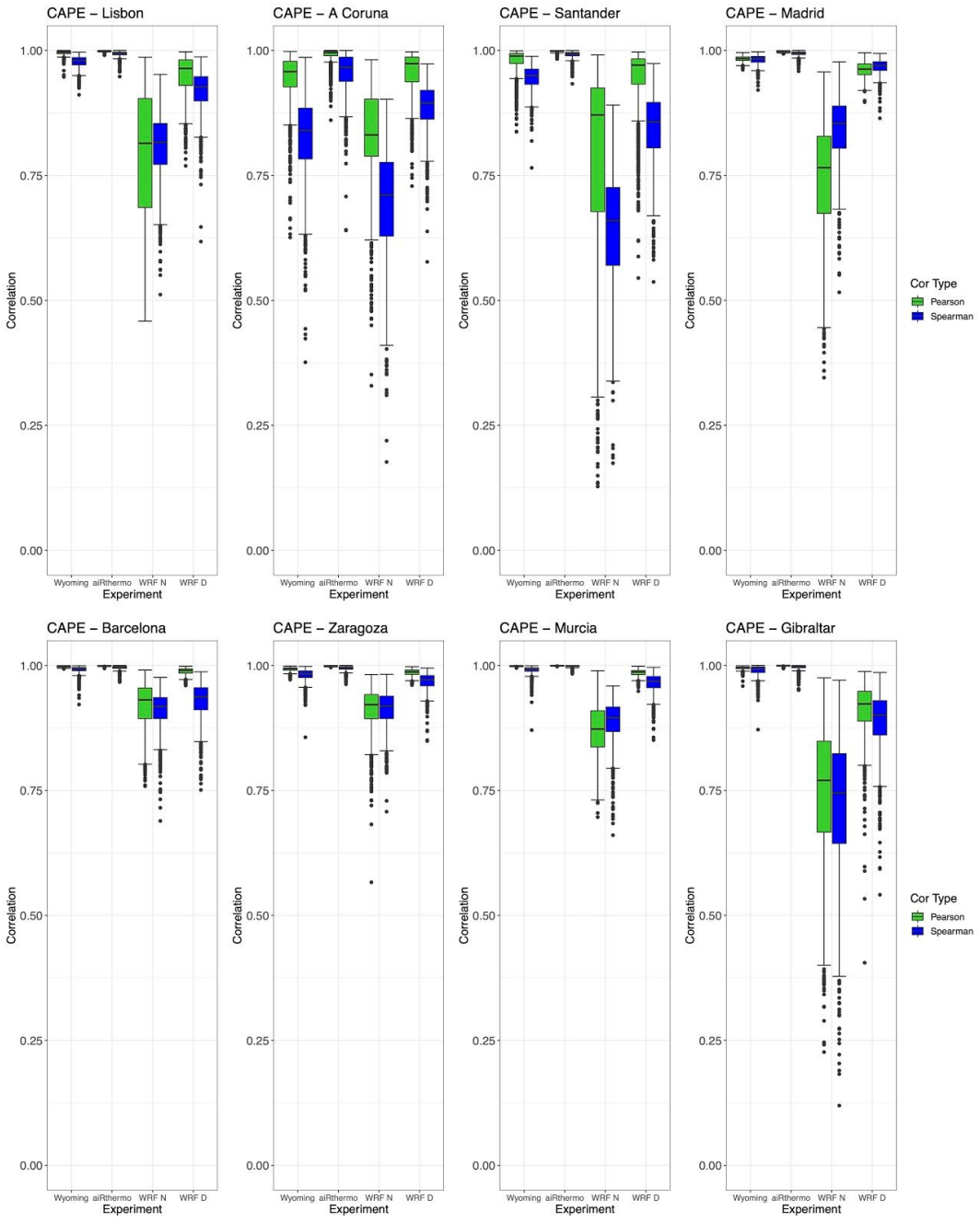

Figure 2: Box and Whiskers for the correlations obtained during the bootstrap for CAPE and calculated by different methods: Pearson in green, and Spearman in blue. IGRA stations are taken as Reference.

>> If we change to Figure 3, the worsening of the correlations is perceptible in A Coruna, Santander, Murcia and Gibraltar. As for CAPE, aiRthermo shows the highest correlations with IGRA dataset, followed by Wyoming, WRF D and WRF N. Particularly for Gibraltar, where the Pearson correlations were similar for both WRF experiments, differences between both of them arise if Spearman correlation is used: in that case, as in the other stations, WRF D obtained better correlations than WRF N. Thus, our conclusions still hold with Spearman correlations, but if we used it, we would lose the nice visual properties of the Taylor diagram and the associated diagnostics using RMSE or fractions of variance, which are also important.

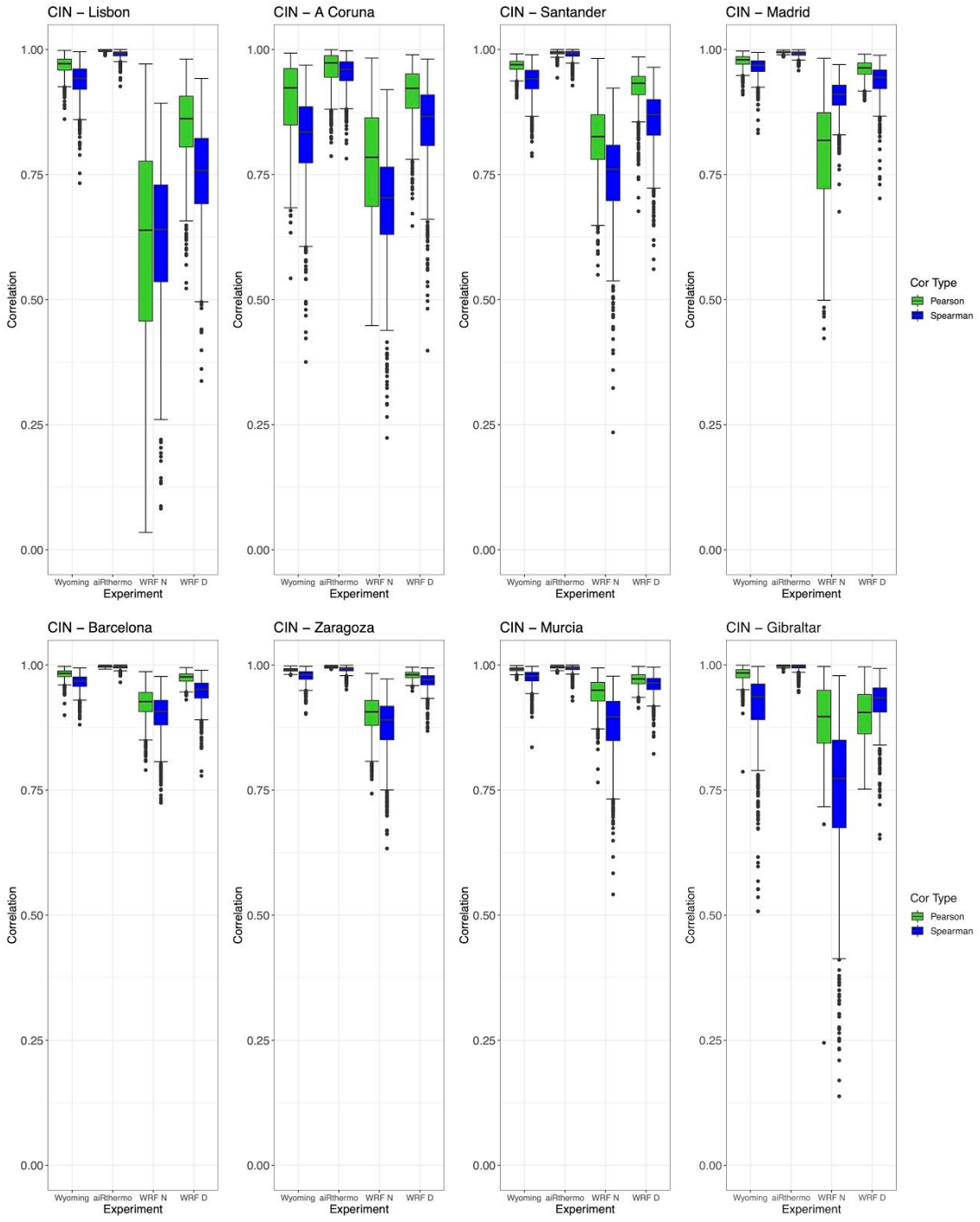

Figure 3: Same as Figure 2 but for CIN.

>> As already said before, Taylor diagrams can only be constructed with the Pearson correlation. Thus, in the new version of the manuscript, both Figures showing the Taylor diagrams for CAPE and CIN will be updated. The Box and Whiskers will be changed to these new versions in order to show also the Spearman correlations.

37. Sect. 2.3.2, last sentence: please delete the statement about precipitation (cf. major point 4).

>> We have edited the sentence as suggested by the reviewer in comment 4 to "These maps show the regions over the IP where the unstable conditions are more prevalent in each season."

38. L188: "by independent locations": Independent in which sense? The locations are not independent right now.

>> We do not understand why the reviewer says that the locations are not independent when they are located in different regions of the Iberian Peninsula, and they are several kilometers far from each other, in different climatic areas and for some fields, such as CAPE with a horizontal scale length smaller than the distance between sites. However, this sentence was edited to highlight these facts.

39. Section 3. Results: To facilitate direct comparison of the subfigures in a panel, it would be very helpful if they have the same axis range.

>> As already stated in Major comment 2, we agree on this comment with the reviewer because having the same scaling in the figures is easier to interpret the results. However, that is not possible in our case because the values show a really large range.

40. L203: Please explain how you selected a model as the "best" model: by the highest correlation coefficient, the lowest rmes, a similar SD, or a combination thereof?

>> As we are using Taylor Diagrams to show the Pearson correlation, RMSEs and standard deviations of each experiment against the reference values, we are following Taylor's suggestions to select the "best" model. As stated by him, "the simulated variables that agree well with observations will lie nearest the point marked 'observed' on the x-axis. These models will have relatively high correlation and low RMSEs, and those lying on the dashed arc will have the correct standard deviation" (Taylor, 2001; Taylor 2005). We will add this information to the new version of the manuscript.

>> ---

>> Taylor, K. E. (2001): Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106(D7), 7183– 7192.

>> Taylor, K. E. (2005). Taylor diagram primer. *Published to web at: https://pcmdi.llnl.gov/staff/taylor/CV/Taylor_diagram_primer.pdf?id=96*

41. L215-216: What is the reason of the small differences between Wyoming and aiRthermo both relying on the same data – in particular for TT which does not involves any assumption about lifting? Why are the differences largest at Murcia?

>> As already stated in minor comment 35, the differences are due to the different methodologies in the calculation of the TT index, mainly due to the Dew Point temperature. In the case of the University of Wyoming, the value is taken directly from the measurement. However, in our case, we do not use that measured value and we calculate it with the Pressure, Temperature and Mixing ratio at 850 hPa. As shown in Figure 1, the values of TT obtained by the University of Wyoming are larger than the ones obtained by aiRthermo. This is observable in Santander, Zaragoza, Barcelona, and particularly in Murcia.

42. L223: Again, are there any reasons why the two stations of Murcia and Barcelona show the largest differences (rmse) compared to the other stations?

>> Again, this is due to the different methodologies followed to calculate TT index. Wyoming University uses the measured Dew Point Temperature, but in aiRthermo we calculate it with the pressure, temperature and mixing ratio at 850mb. There may exist different methods to estimate saturation pressure of water vapour implied in the calculation of Td. See minor comment 35 and 41.

43. L231: "...small differences in initial conditions..."; can you be more specific here (also with regard to TT, as already alluded above)?

>> As replied in our previous comments, the differences are due to the fact that Wyoming University uses the measured Dew Point temperatures to calculate its indices, while our methodology is only based on pressure, temperature and mixing ratio. In the case of aiRthermo, we only used those measurements and we do not include Dew Point Temperature in any case.

>> For the calculation of CAPE and CIN, the Lifted Condensation Level (LCL), the Level of Free Convection (LFC) or the Equilibrium Level (EL) must be calculated. Depending on the methodology used (the number of low levels averaged for the initial parcel, the definition of the saturated pressure as a function of temperature, truncation errors due to the number of digits used in the ascii files storing the soundings to name three examples), the location of these levels can vary and this can trigger differences in CAPE and CIN. Everything starts with the calculation of the LCL, and that is calculated using the Dew point temperature. Small differences in those values can trigger differences in the values of CAPE and CIN.

>> Even if differences are expected, thanks to Figures 6 and 7 of the current version of the manuscript, we can see that the distribution is similar for both Reference (Wyoming data) and aiRthermo in all the stations. This is also applicable to the TT index, as shown in Figure 1 of this reply, and Figure 5 of the manuscript.

44. Figure 3 (CAPE) shows very large differences of the standard deviation between the different models and for some of the stations. Any idea on that?

>> The only difference in the configuration of both WRF experiments is the data assimilation in the D experiment. Thus, it is clear that this should be the reason.

45. L250: Could you be more specific?

>> This sentence has been edited to highlight the ability of the data assimilation to produce more reliable results (similar to those derived from measured data).

46. L254: “N tends to overestimate the variability in every season and for most of the stations...” Why?

>> Again, the only differences between WRF experiments is the data assimilation scheme included in D. Thus, the data assimilation corrects the temperature, pressure or/and mixing ratio from the model.

47. L255: “..presents the largest values during winter, which agrees with the fact that the northern and northwestern IP receives greats amount of rain during that season”. Is winter rainfall really dominated by convective precipitation? I cannot find any statements in the cited literature. Which of the Atmospheric patterns AP1-19 defined by Romero et al., 1999 are convective patterns? Rodriguez-Puebla et al., 1998, considers only the relation to teleconnection patterns, but did not classify precipitation.

>> In that sentence we are not referring only to convective precipitation, as can be inferred from the cited papers, which are also discussed in the Introduction in order to present the seasonal regimens of precipitation observed in the IP. As stated in that line, we only highlight that the largest TT values are located in the regions where more precipitation is measured during winter, without any comment related to convective precipitation.

>> In order to avoid wrong interpretations, this line have been rewritten in the new version of the manuscript.

48. L255-260 and Fig. 5: Why does Lisbon show higher TT values in winter than in summer? You may also mention that the differences between the models at Lisbon, La Coruna, and Santander are larger in winter compared to summer (why?).

>> Figures 4, 5 and 6 show the seasonal mean maps (winter and summer only) for the variables involved in the calculation of TT index, that is, temperature and dew point temperature at 850 hPa and temperature at 500 hPa respectively.

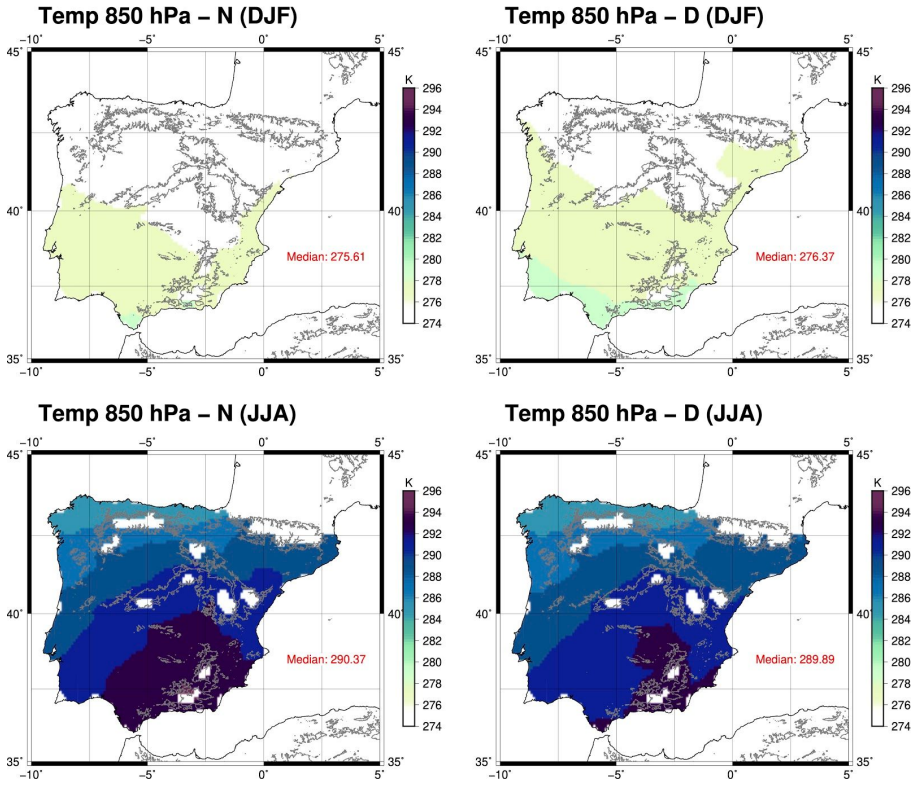

Figure 4: Spatial distribution of mean Temperature at 850 hPa for period 2010-2014 over the IP as computed from N (first column) and D (second column) for winter and summer. The median value (K) is in the bottom right corner of the plots.

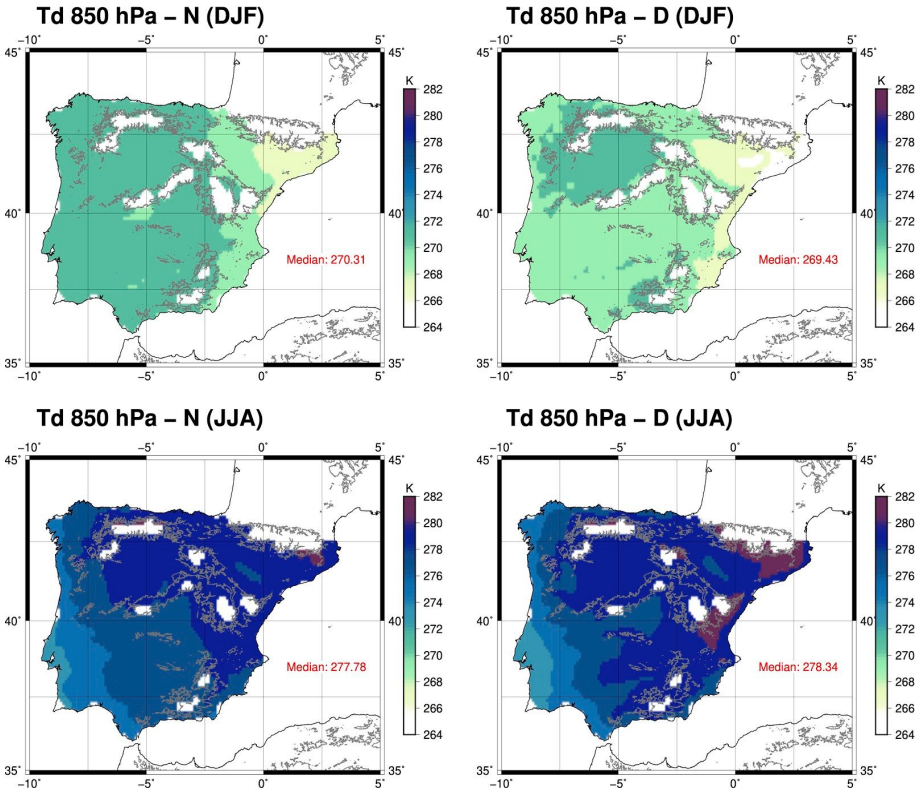

Figure 5: Same as Figure 4 but for dew point temperature at 850 hPa.

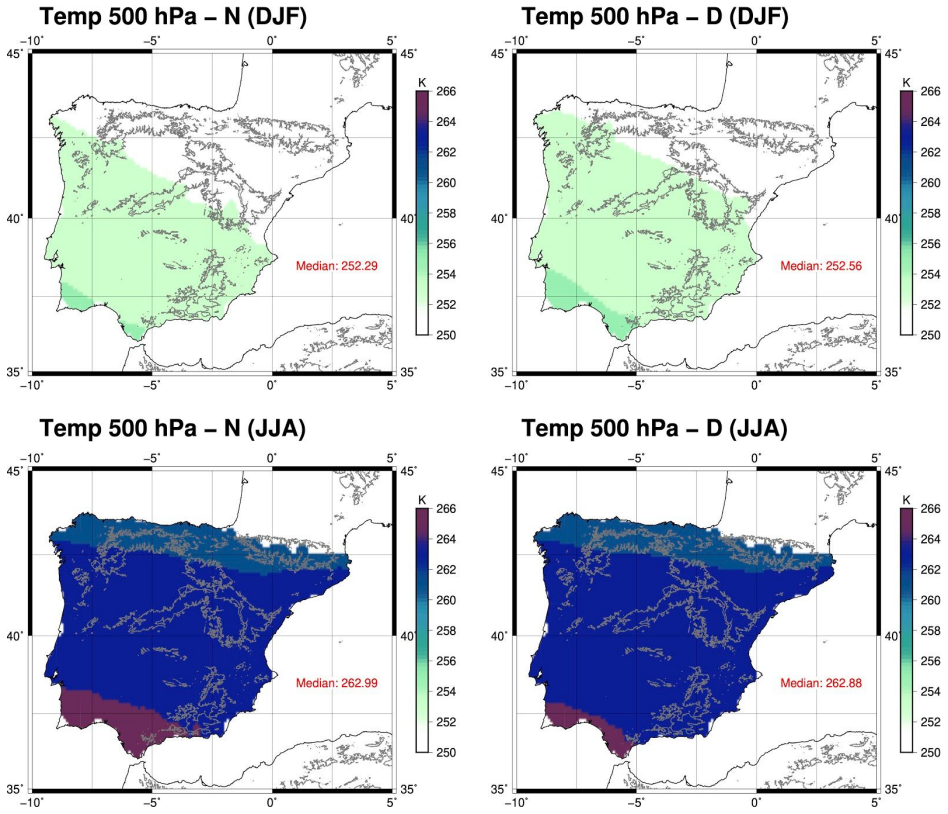

Figure 6: Same as Figures 4 and 5 but for temperature at 500 hPa.

>> Taking into account the information from these figures, the shorter values of TT obtained in Lisbon during summer than in winter are due to the the shorter Td values obtained in that area in summer. Particularly, the values of temperature at 850 and 500 hPa increase in that region from winter to summer (about 15 and 10 degrees respectively), while the dew point temperature is only a few degrees larger. That produces the reduction of the TT values from winter to summer near Lisbon.

>> These Figures also depict the differences between both WRF experiments. In Figure 4, we can see that the data assimilation increases the temperatures over the IP in winter, while it reduces them in summer (particularly in the southeastern corner of the IP). For the dew point temperature, the reverse is observed in Figure 5: the dew point temperatures are reduced in winter, while they are increased in summer (with the exception of the western facade of the IP where they are reduced). In Figure 6 we can see that the temperatures at 500 hPa are slightly increased in winter and slightly reduced in summer. These differences are the ones shortening the differences between both WRF simulations from winter to summer.

>> We do not think that these figures should be included in the final manuscript. Thus, a short summary of these results will be included in the text and the figures will be provided as supplementary materials.

49. L261-262: Again I miss a reference that shows not only total precipitation, but a classification among the types (stratiform/convective).

>> As already said in comment 47, in those lines we are not restricting our results to convective precipitation, and we only highlight the fact that these regions are the ones obtaining more precipitation in that season.

>> The line will be rewritten to avoid misunderstandings in the new version of the paper.

50. L268-270: Why does D overestimate CAPE at most stations in spring? And why does the N experiment overestimate CAPE in winter and underestimate it in summer?

>> Figure 7 shows the median of vertical profiles of virtual temperature for the soundings and the lifted parcels until 550 hPa in spring. The dashed lines represent the 5 and 95 percentiles. If we focus only on the soundings from IGRA, WRF D and WRF N, we can see that in general, D is closer to the observed virtual temperature (particularly in A Coruna, Santander, Barcelona and Zaragoza). Additionally, both WRF experiments tend to warmer conditions between 800-750 and colder near surface (until 900 hPa). If we switch to the lifted trajectories, these are warmer for D and colder for N in most of the stations. Additionally, N tends to cross the sounding later than D (e.g. Lisbon, Santander or Gibraltar).

>> Thus, both WRF simulations overestimate CAPE during this season due to the differences in the virtual temperature in lower levels (colder near surface and warmer near 800 hPA compared to measured data). In combination with the fact that the lifted trajectories for D are slightly warmer than the observed ones and N, this experiment overestimated CAPE in most of the stations during that season.

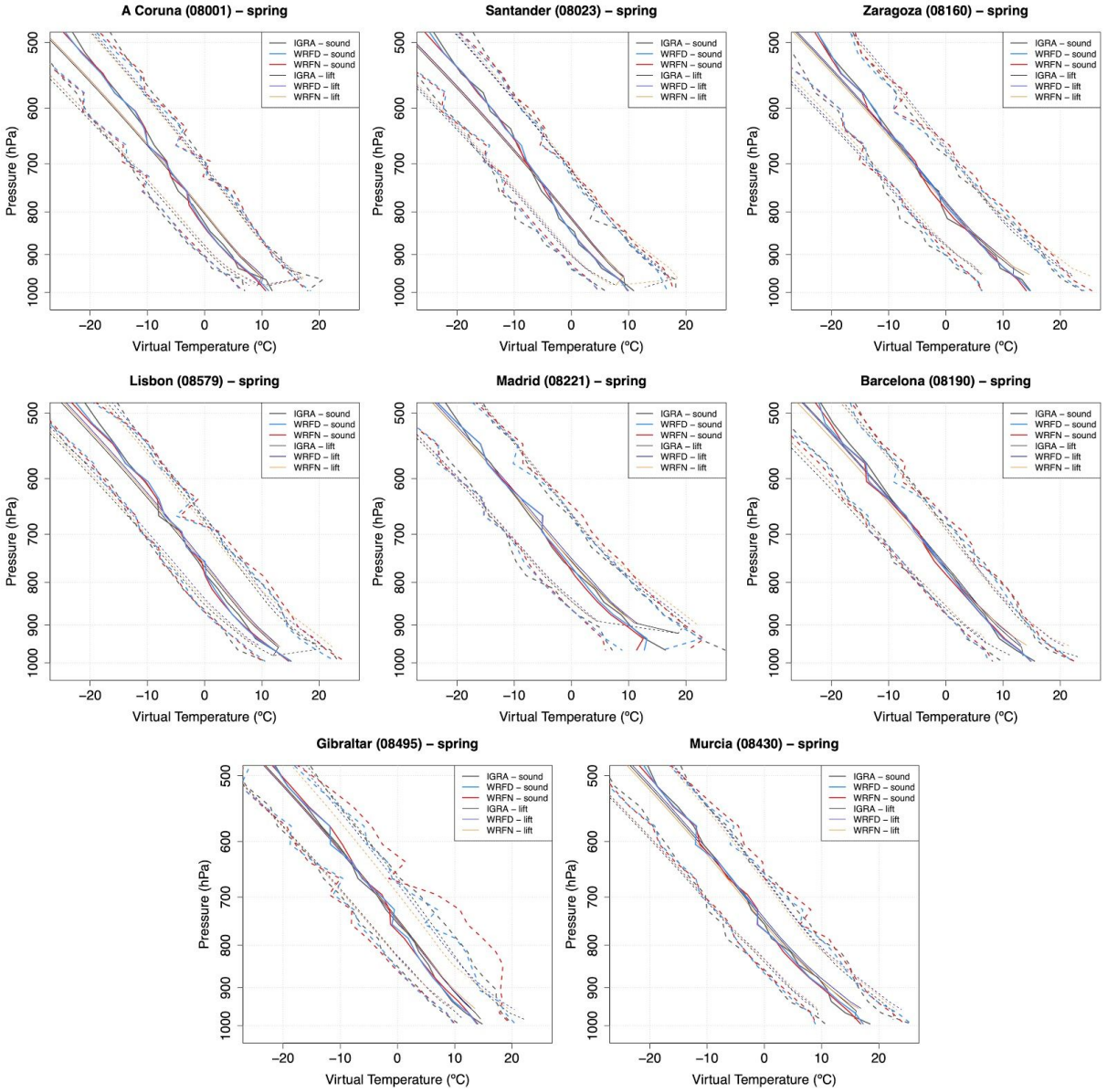

Figure 7: Vertical profiles of virtual temperature for the sounding levels and for the lifted parcel during spring. The dashed lines represent the 5 and 95 percentiles, and the solid lines the median.

>> Figure 8 shows the results for Winter. If we focus on the soundings, we can see that in most of the stations D is much more similar to IGRA than N. During these season, the soundings tend to be colder in low levels (below 800 hPa) for N compared to IGRA. In the case of the lifted trajectories, these are warmer for N. Thus, the combination of these two factors (colder soundings and warmer lifted trajectories) cause the overestimation of CAPE for the N experiment. This is well observed in Lisbon, A Coruna and Santander.

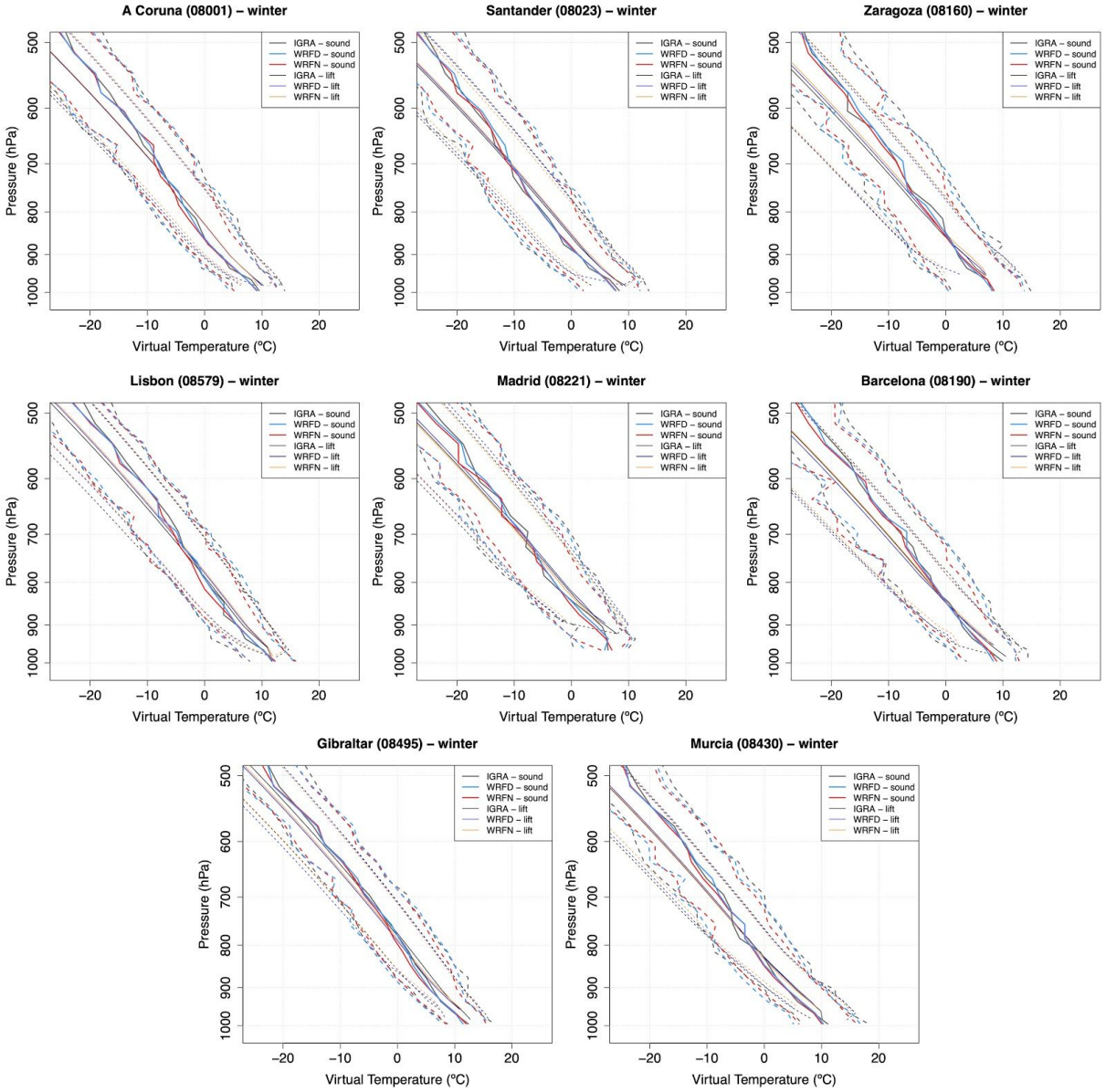

Figure 8: Same as Figure 7 but for winter.

>> If we change to summer, Figure 9 shows the corresponding trajectories and soundings. In this season, N shows an overestimation of CAPE, which is clearly observable in Barcelona, Murcia and Gibraltar. In these three stations, N shows warmer sounding levels that IGRA, which produces that the lifted trajectory crosses earlier than the other experiments the sounding, and consequently, underestimating the CAPE.

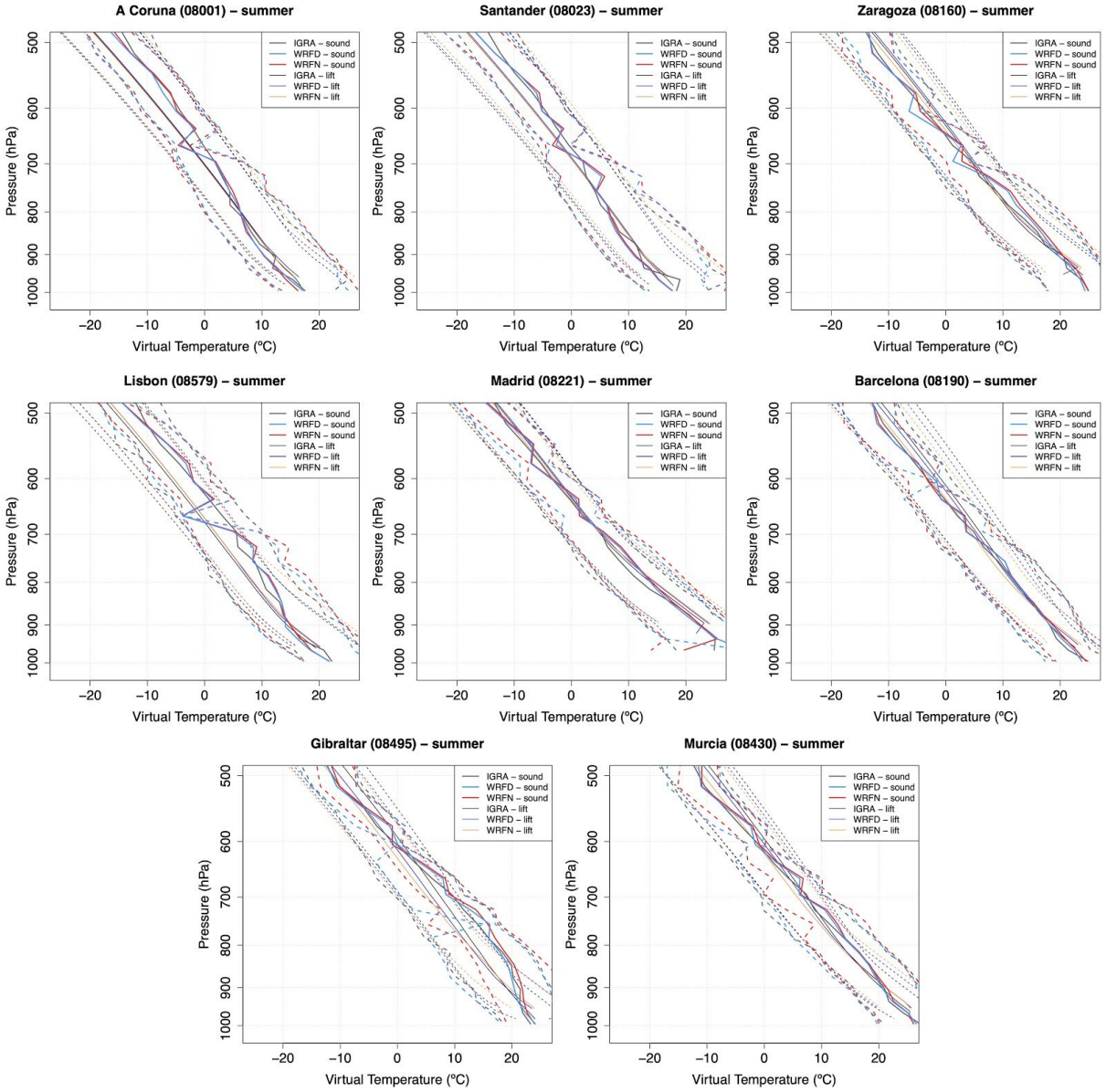

Figure 9: Same as Figures 7 and 8 but for summer.

51. Figs. 5/6: At Barcelona, Zaragoza, and Murcia, CAPE is highest in summer, whereas TT reaches highest values in spring at these stations. What is the reason of the obvious discrepancy between CAPE and TT?

>> If we focus on those stations highlighted by the reviewer, it is clear that it seems to be a discrepancy between those indices (even if the differences between the TT values between spring and summer are below 2 degrees). However, the reviewer forgets that these values of CAPE (and also CIN - Fig 7) are calculated for the entire series of 12 hourly values obtained in each station during 2010-2014. Thus, these values are not restricted to highly convective events. Consequently, Fig 6 and 7 must be compared to the values of TT in combination.

>> If we do so, we can see that the not extremely high values of CAPE observed in Barcelona, Zaragoza and Murcia during spring are contrasted with the highest values of CIN of the season in those stations. In contrast, during summer, extremely high values of CAPE are observed in those stations (medians over 200 J/kg for Barcelona and Murcia, but they can reach values over 700 J/kg), but CIN is not comparable to those values of CAPE (below 150 for Barcelona and Zaragoza, around 200 for Murcia). Thus, Barcelona and Zaragoza present unstable conditions in summer, but not in spring.

>> Since CAPE and CIN are dependent on the entire profile of the atmosphere, CAPE and CIN should be more reliable than TT.

52. L281 and following: As already mentioned above (see major comment 5), CIN is relevant for convection only in combination with CAPE (An example: imagine a day with zero CAPE and zero CIN; another day with CAPE = 3000 J/kg and CIN = 300 J/kg. None of the days would have the right conditions for deep moist convection to occur. The average of the two days would give CAPE = 1500 J/kg and CIN = 150 J/kg. Fair values for DMC). You could simply fix that by considering CIN only on days for CAPE in excess of 50 or 100 J/kg.

>> As already stated in major comment 5, that is not the objective of our paper. The objective of this paper is to evaluate the ability of WRF simulations (including or not the 3DVAR data assimilation step) to produce reliables values of TT, CAPE and CIN over the Iberian Peninsula. Once that has been evaluated by means of Taylor diagrams, we want to show the distribution of the values of the complete time series from 2010 to 2014 for each index. We are not trying to make any prognosis or diagnosis of convective events from the data we prepared.

53. L305: "...lowest values are observed near the coastal valleys..." why?

>> Figures 10, 11 and 12 show the seasonal mean maps for 00 and 12 UTC in winter and summer for temperatures and dew point temperatures at 850 hPa, and temperature at 500 hPa. According to these results, the lowest values of TT observed near the coastal valleys are originated due to the low values observed in those regions for dew point temperature, which at the same time is originated by the low mixing ratio values in those regions. The low values are observed at 00 UTC, and they are even lower at 12 UTC. As a result, the TT values are low in those coastal regions, independently of the facade of the Iberian Peninsula. This information was added to the new version of the manuscript, but not the Figures as they look similar to the means for winter and summer included in comment 48 (which will be included to supplementary materials).

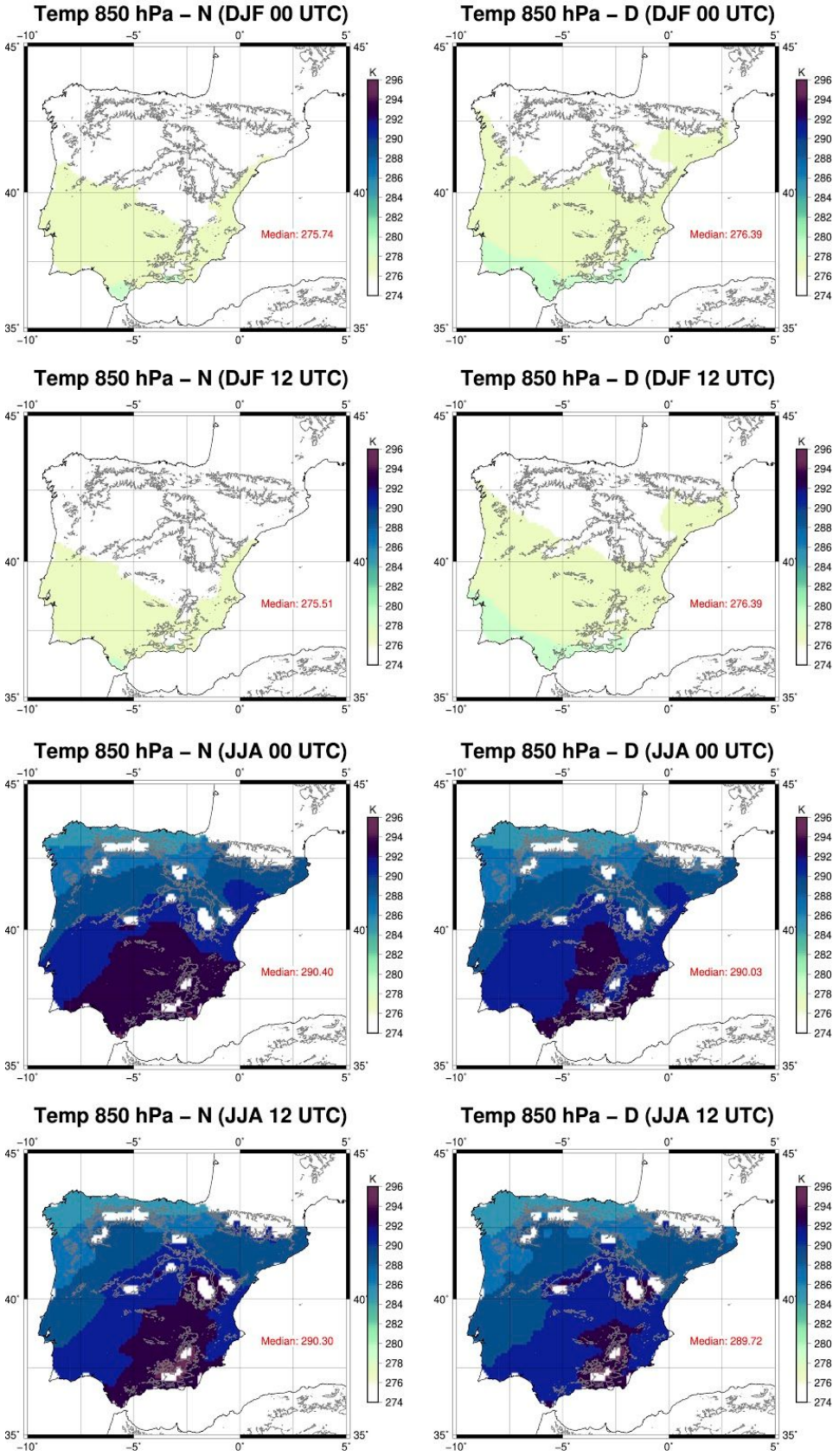

Figure 10: Spatial distribution of mean Temperature at 850 hPa for period 2010-2014 over the IP as computed from N (first column) and D (second column) for winter and summer at 00 and 12 UTC. The median value (K) is in the bottom right corner of the plots.

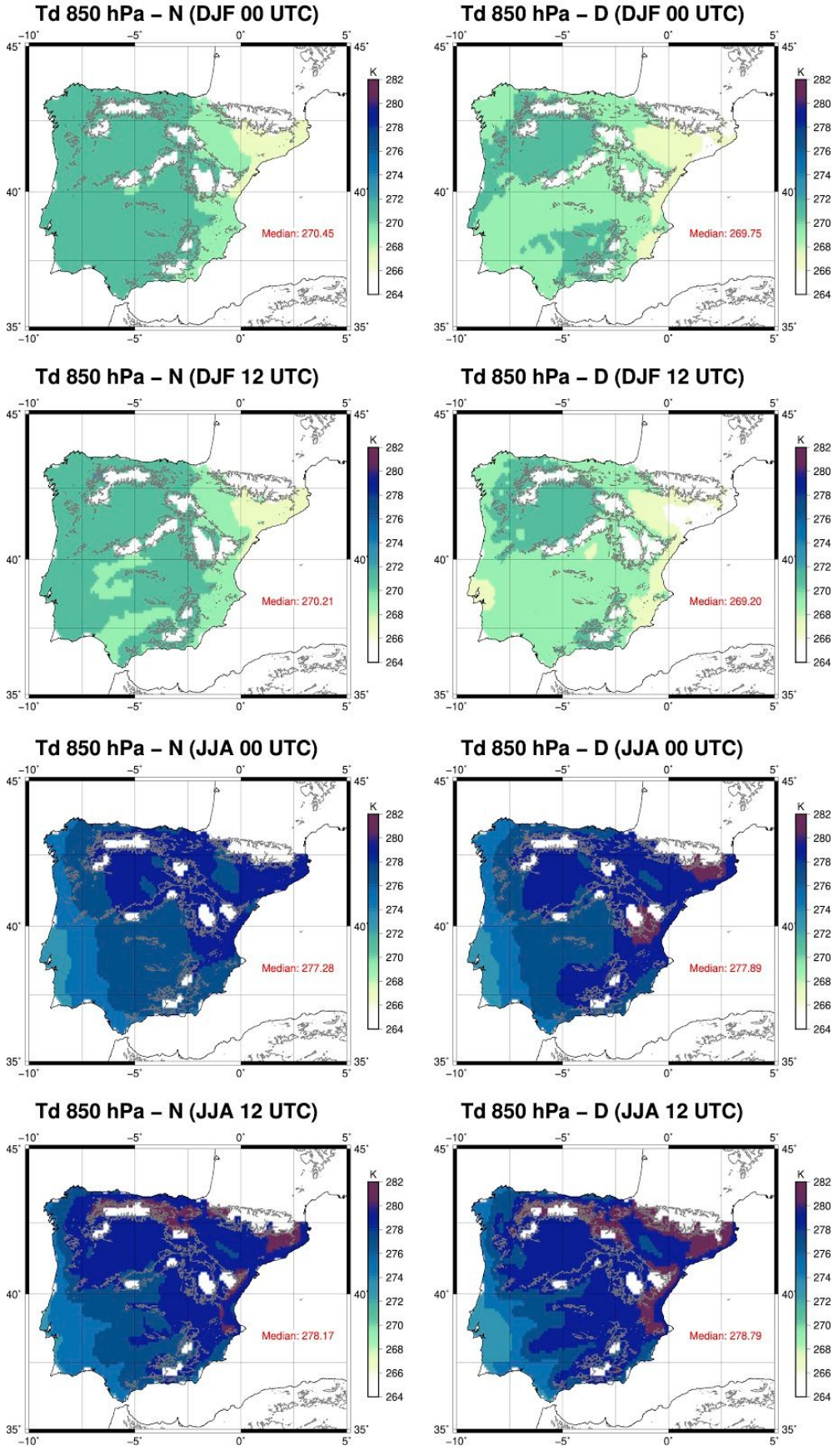

Figure 11: Same as Figure 10 but for dew point temperature at 850 hPa.

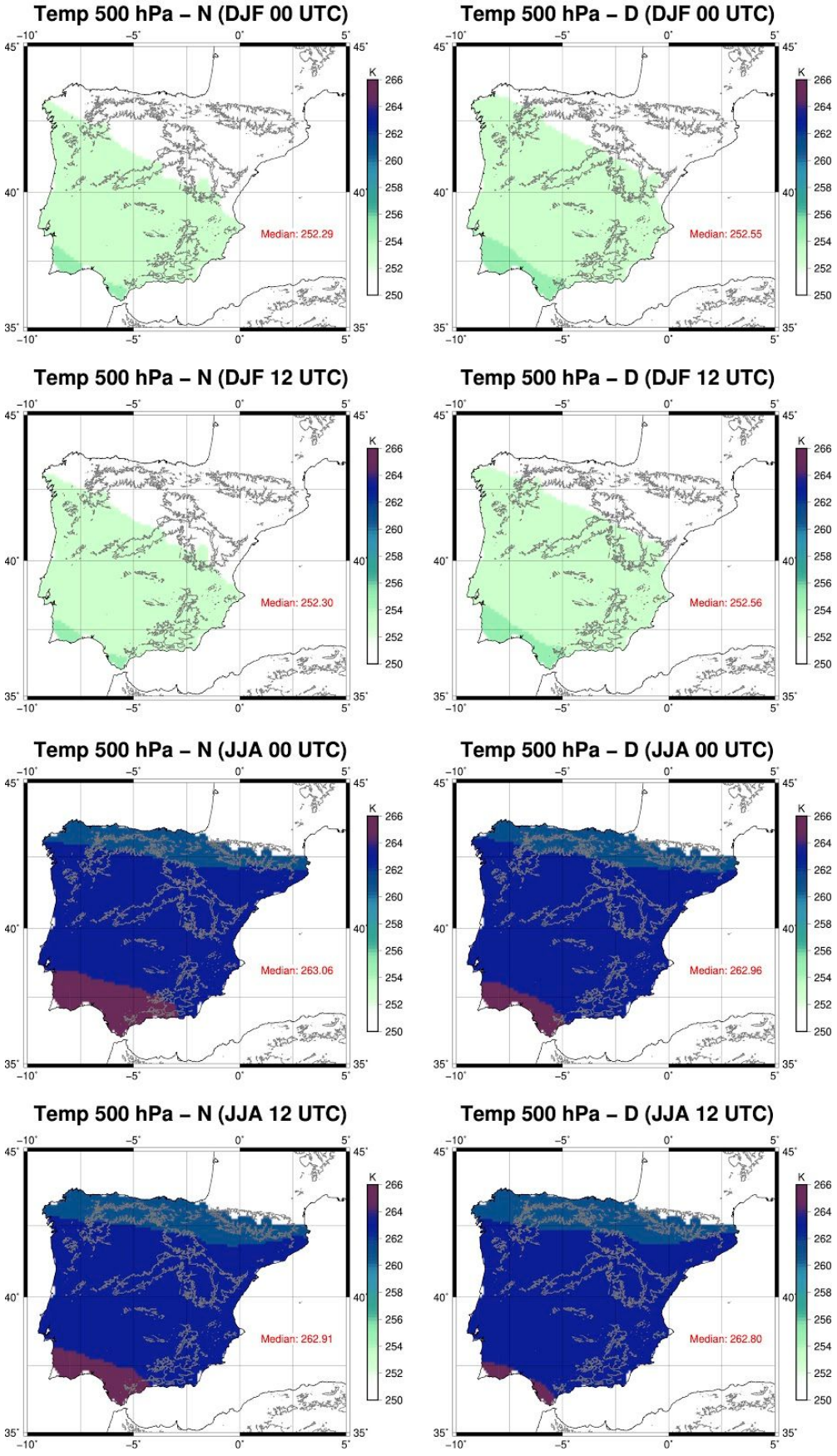

Figure 12: Same as Figures 10 and 11 but for temperature at 500 hPa.

54. Figure 8/9: The spatial distribution of TT and CAPE in most of the cases is contrary, i.e. regions with higher CAPE have lower TT values and vice versa. Any explanation of this apparent contradiction?

>> The apparent discrepancy highlighted by the reviewer is only observed in winter near Lisbon and the western facade of the IP. This is not observed in summer, where the highly unstable areas are mainly observed towards the Mediterranean coast in both TT and CAPE.

>> However, it must be taken into account that our results for CAPE and CIN are not restricted to highly convective events, and they represent the mean values computed during 2010-2014. As can be seen in Figures 13 and 14 (A Coruna and Gibraltar as an example), TT and CAPE are related to atmospheric instability, but they are not related through a simple linear relationship (R2 below 0.2 for all the stations and seasons), particularly for stable or neutral atmospheres. Since we are showing the results corresponding to all the observations, the relationship does not need to be simple. This is expected, since TT is a diagnostic computed from discrete levels and CAPE and CIN involve the vertical integral along the atmosphere.

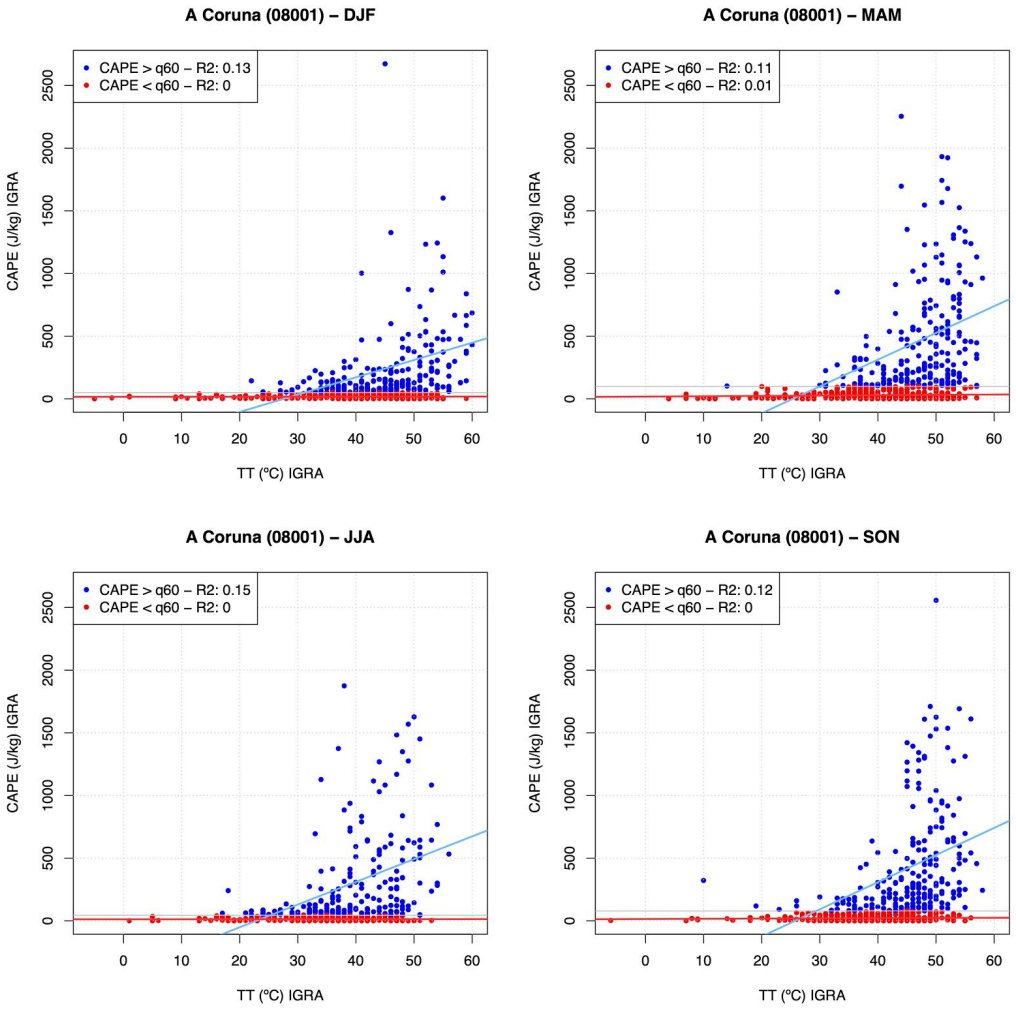

Figure 13: Scatterplots for the values of CAPE and TT as included in IGRA for A Coruna. The values of CAPE over the 60th percentile are in blue, and the values below that value are in red. The value of the 60th percentile is marked with a grey line, and the linear models are also included with the corresponding colors.

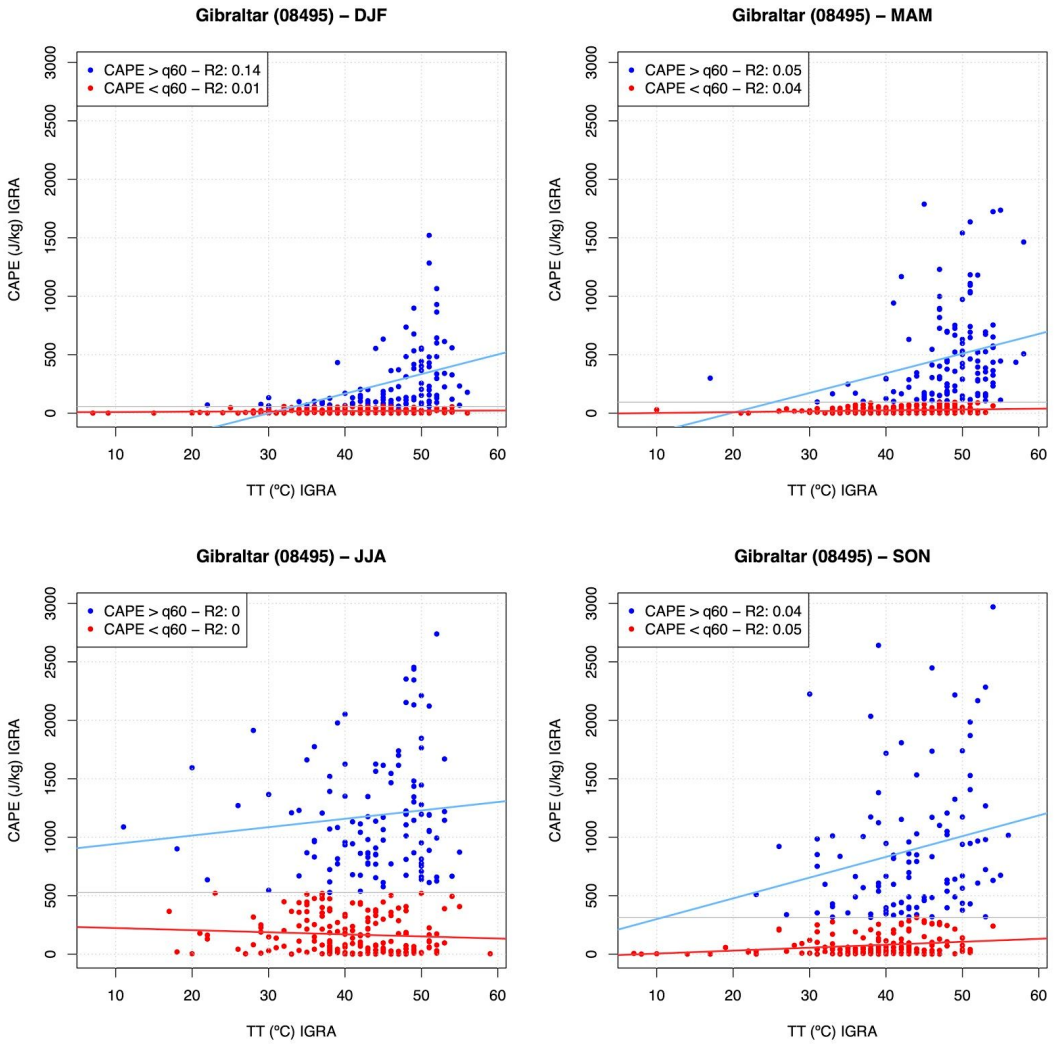

Figure 14: Same as Figure 13, but for Gibraltar.

>> This apparent discrepancy was addressed in the new version of the manuscript, and the above presented details were included in the new version of the manuscript..

55. L326 and following: See major comment 5 and minor 52.

>> As already stated in those comments made by the reviewer, restricting the values of CIN and not including the entire time series for period 2010-2014 is not part of the objective of our paper. We are not restricting the evaluation of these instability indices to the extreme events, and we are evaluating the performance of the model at simulating them. In this case, in order to show the different patterns obtained by each WRF experiment, the mean values of the entire period where chosen.

56. L336-345: The relation to “dynamics” does not fit here as the paper solely has a thermodynamical perspective. Be careful with the relation between convective conditions and precipitation.

>> The term “dynamics” was used only to introduce the fact that two different patterns are observed between winter and summer. In any case, it was not related at all to the relationship between convective conditions and precipitation. Thus, those lines were edited as suggested by the reviewer.

Edits:

1. L5: explain “IP“ at first use; the same applies to “SD” in L8

2. L6-7 “measured variables at different pressure levels” may be replaced by “vertical temperature and moisture profiles”.

3. L9 methodologies methods

4. L11: “an air parcel’s vertical evolution” “a lifted air parcel”

5. L13: “...reference values” of which quantity?

6. L15: “**in** the Mediterranean coast” “**over** the...”; never use in the coast, in the station etc.

7. L16: “The chances of developing thunderstorms in those areas at 12 UTC is much higher than at 00 UTC”. This generally applies to convection. Delete.

8. L18: Murcia: Don’t mention a location in the abstract that is not well known.

9. L22: “in the planet”?? “of the planet”

10. L39 “alone and it should be...” “alone, but should be...”

11. L39: There are much more lifting mechanisms, thus include “, for example, by orography, ...”

12. L48: **On** the global scale

13. L50: “convective storms develop for lower values...”

14. L62: “CAPE presents” “CAPE shows a high...”

15. L72-73: use the plural: hailstorms and thunderstorms

16. L74: “similar to those observed in Europe..” Where in Europe? Is the IP not part of Europe? What is meant by “dispersion of the values”?

17. L100: exercises experiments

18. P4: please explain all abbreviations at first use

19. L109-110: “.., and they use 51 levels” “and with 51 levels”

20. L111: form from

21. LL123: Both simulations, **N and D**, ...” (note that you refer to another simulation in the previous section).

22. L136: either delete “..where they only measure it at...” or rewrite this sentence

23. L141: suggestion: “Additionally, vertical profiles of pressure, temperature, and mixing ratio obtained...”

24. L142: delete “indices”

25. L147: “This pattern...” which pattern do you refer to? What is meant by intensity?

26. L154: better use WRF’s original eta levels instead of models’s to avoid any misunderstanding that you quantified CAPE/CIN only from the 20 ERA-Interim levels.

27. L187: “...of the error/differences due to the different methods applied”

28. L203: be careful with the wording, aiRthermo is not an experiment but simple a quantification.

29. L231: trigger result in

30. L233: “...between **the** experiments”

31. L255: “A Coruna and Santander...”

32. L273: What do you mean by “some stations are more important than others...”?

33. L295-296: “Figure 8 shows...” and “The heterogeneity...” reformulate these two sentences

34. L301: important higher

35. L317: “...the unstable area **air mass** is found...” An area cannot be unstable

36. L318: For readers outside of the IP, can you give a hint about the location of Tagus and Guadalquivir rivers? And later, L331: the Ebro basin?

37. L320: “On the contrary **Compared** to what...” Or to which contradiction you refer?

38. L321-322: “where tunderstorms can be developed” “the area with higher CAPE values...”

The edits suggested by the reviewer were applied to the new version of the manuscript.