

Dear Editor,

Please find enclosed the revised version of the manuscript entitled “Changes in the simulation of instability indices over the Iberian Peninsula due to the use of 3DVAR data assimilation” by S. J. González-Rojí, S. Carreno-Madinabeitia, J. Sáenz and G. Ibarra-Berastegi, that we resubmit to the journal *Hydrology and Earth System Sciences*.

All the major points raised by both reviewers (and by Dr. Klemens Barfus) during the open review have thoroughly been addressed in the current version of the manuscript. Additionally, 6 extra figures were included as annexes in the new version in order to support some new statements made in the manuscript and to clarify some points that were raised by Reviewer#2. Attached to this cover letter you will find the tracked changes version of it, where all the thorough modifications are highlighted.

As explained in the detailed response to the reviewers published online, we consider that we successfully addressed all the points raised by the reviewers and, as such, we hope that the manuscript could be accepted this time.

Yours faithfully,
Santos J. González-Rojí

Comment made by: Klemens Barfus (klemens.barfus@tu-dresden.de)

Received and published: 30 March 2020

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

Dear authors,

I find this paper very interesting and well written. Nevertheless, from my point of view clarification about the vertical levels of ERA-Interim used as input to WRF is needed.

>> Thanks for your kind words and we appreciate these insightful comments.

Authors write that they used 20 levels without providing further details. In González-Rojí et al, 2018 we find the information that levels range from 5 hPa to 1000 hPa. I conclude from this that authors work with data on pressure levels available from the ECMWF servers.

Counting the available pressure levels between 5 hPa and 1000 hPa from ECMWF, I find 34 levels. When data on model levels are used, 60 levels are available (terrain following and thus with different pressure levels at each grid point).

>> Yes, we feed the WRF model with **analyses** of temperature, relative humidity, both wind components and geopotential at 20 pressure levels downloaded from the MARS repository. The exact pressure levels are: 5, 10, 20, 30, 50, 70, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 850, 900, 925, 950, 1000 hPa.

>> The data in the original model levels from ERA-Interim are represented in spherical harmonics, and we feed the inputs to WRF model in a regular longitude/latitude grid. The reason is that all the variables needed by WRF are obtained only in that regular grid. Only temperature is available in the original model levels from ERA-Interim (see Table 3 from Berrisford et al., 2011). Those are the reasons that we didn't follow the path suggested by Dr. Barfurs.

Berrisford, P., Dee, D., Poli, P., Brugge, R., Fielding, K., Fuentes, M., Kallberg, P., Kobayashi, S., Uppala, S. and Simmons, A., 2011. The ERA-Interim archive, version 2.0. <https://www.ecmwf.int/en/elibrary/8173-era-interim-archive>

The calculation of CAPE and CIN is sensitive to the vertical resolution of the data.

If authors used pressure level data, information about the used levels and why they did not use all available pressure levels is needed. Furthermore, I would like to read why authors did not use model level data, since they provide much more information about the temperature and humidity profiles especially in regions with high topography.

If authors used model level data, I also would like to read information about the used levels and why not the full set of available levels have been used. From my point of view some information provided by the assimilation could already be enclosed when using all available model levels.

>> We know that both CAPE and CIN are sensitive to the number of vertical levels used in their calculation, but concerning ERA-Interim, we have not calculated the values of these indices. Neither in the original model data nor in the downloaded pressure levels. We have not used in our paper the CAPE values available in the forecast stream (not the analyses stream) of ERA-Interim (see Table 8 from Berrisford et al., 2011). Since the objectives of our paper do not include a comparison of WRF versus ERA-Interim, we do not see necessary to add this information.

>> Regarding the indices calculated with the measured data from the radiosondes from Wyoming (aiRthermo in the manuscript), all the available pressure levels in the soundings were used for the calculation of TT, CAPE and CIN.

>> In the case of both WRF simulations, the calculations were done using all the available eta model levels from our configuration. As stated in the manuscript, 51 vertical η levels up to 2000 hPa (value at the top of the atmosphere) are available.

>> All this information about the calculation of the indices with the pressure levels available for each option (aiRthermo and both WRF simulations) is already stated in the subsection 2.3.1. However, in the next version of the manuscript, all this information will be stated in a clearer way.

Berrisford, P., Dee, D., Poli, P., Brugge, R., Fielding, K., Fuentes, M., Kallberg, P., Kobayashi, S., Uppala, S. and Simmons, A., 2011. The ERA-Interim archive, version 2.0. <https://www.ecmwf.int/en/elibrary/8173-era-interim-archive>

A minor issue: as far as I know, there is no quality control for the Wyoming radiosondes. Authors should provide information about their own QC routines. Waiving the high vertical resolution of Wyoming radiosondes (from my point of view not a good idea when analyzing CAPE and CIN), IGRA quality controlled radiosondes could be used.

>> The radiosondes from Wyoming University were already used for the validation of the precipitable water from our both experiments in González-Rojí et al. 2018, and none of the values were taken as erroneous. For the continuity of the study related with both simulations, we decided to use the same radiosondes for the calculation of the indices. Additionally, since the values of CAPE are very sensitive to methodological factors, such as the computation of the initial parcel or the vertical spacing in pressure levels, we have estimated CAPE and CIN 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).

>> In order to validate the CAPE, CIN and TT indices calculated by Wyoming University, they were compared with the ones calculated by IGRA as suggested by Dr. Barfurs. 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.

>> Figure 1 shows by means of Taylor diagrams the comparison of previous results included in the paper against the TT index calculated from the IGRA soundings (all levels), which will be chosen as the reference. It is clearly seen that the closest points are always without almost no exception the green ones (values of TT reported by the Wyoming archive), and the yellow ones (those computed by our package aiRthermo from the pressure levels from Wyoming archives) with the only exception of Murcia. The next best determination of the TT index is achieved by the D experiment (WRF run including data assimilation, blue points) and, finally, the red points corresponding to the N simulation (WRF run without data assimilation) shows the worst agreement in all cases. This implies that the use of IGRA data as the reference instead of the Wyoming soundings does not change the conclusions of our study (even the exception for Murcia was observed in our previous results). It also shows that the error due to homogeneity in such a short interval of time (five years) is very small (at least, in this observational record).

>> Figure 2 shows a similar result for CAPE. The estimation of CAPE from Wyoming and the one estimated with aiRthermo using pressure levels from Wyoming soundings are always quite close to the value estimated from IGRA soundings (RMSE smaller than 20 J/kg in all cases). As in the case of TT, the main results of

our paper are not affected by switching the reference dataset from Wyoming to IGRA: the D experiment shows better agreement with observed CAPE than the N experiment.

>> Figure 3 shows the same result for CIN. The RMSE of CIN values computed from Wyoming soundings and IGRA soundings is also small (smaller than 20 J/kg). With the only exception of Gibraltar, in which N and D behave exactly the same (this was also observed in the paper), D always produces better results than the N simulation.

>> Thus, we find that our results are robust to the selection of the observational dataset. However, the figures that we show here will be incorporated to the final version of the paper, since we feel this comment leads to a better paper.

González-Rojí, S.J., Sáenz, J., Ibarra-Berastegi, G. and 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), pp.708-729.

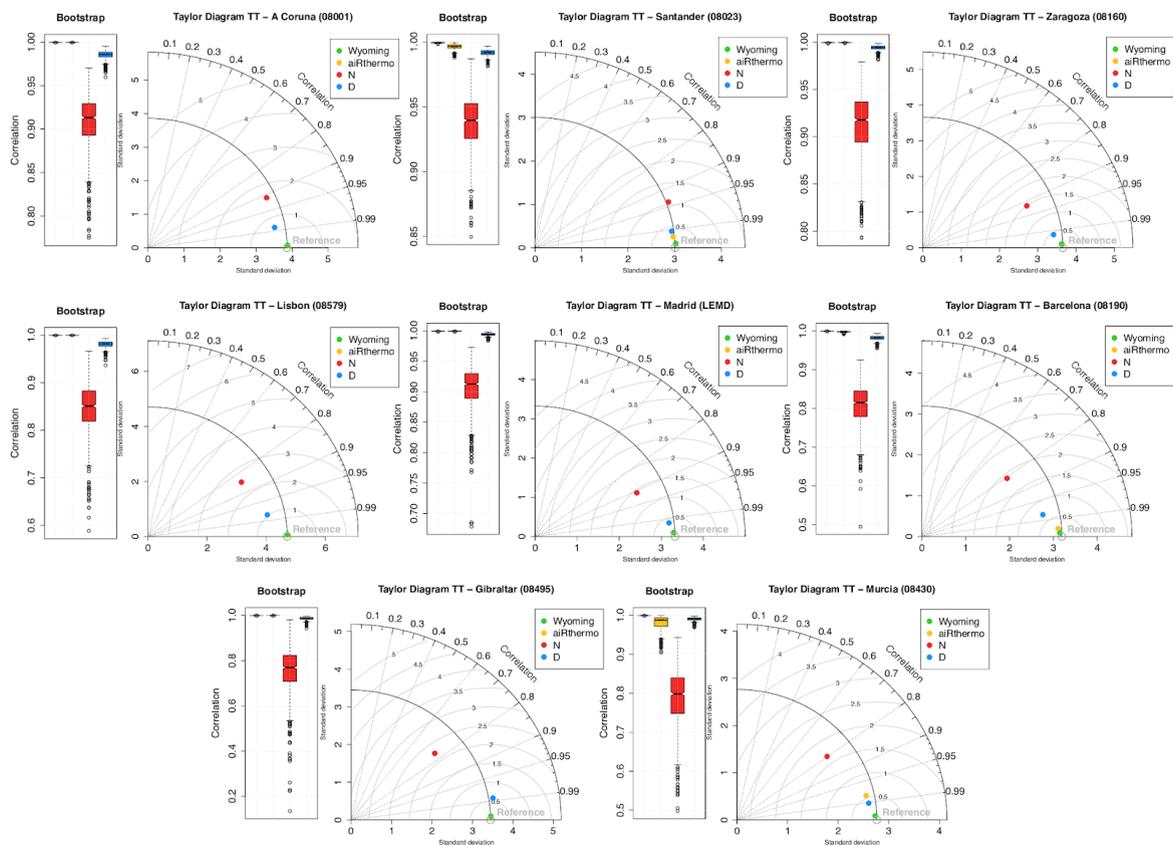


Fig.1 Taylor diagrams of TT index if the values from IGRA are taken as reference.

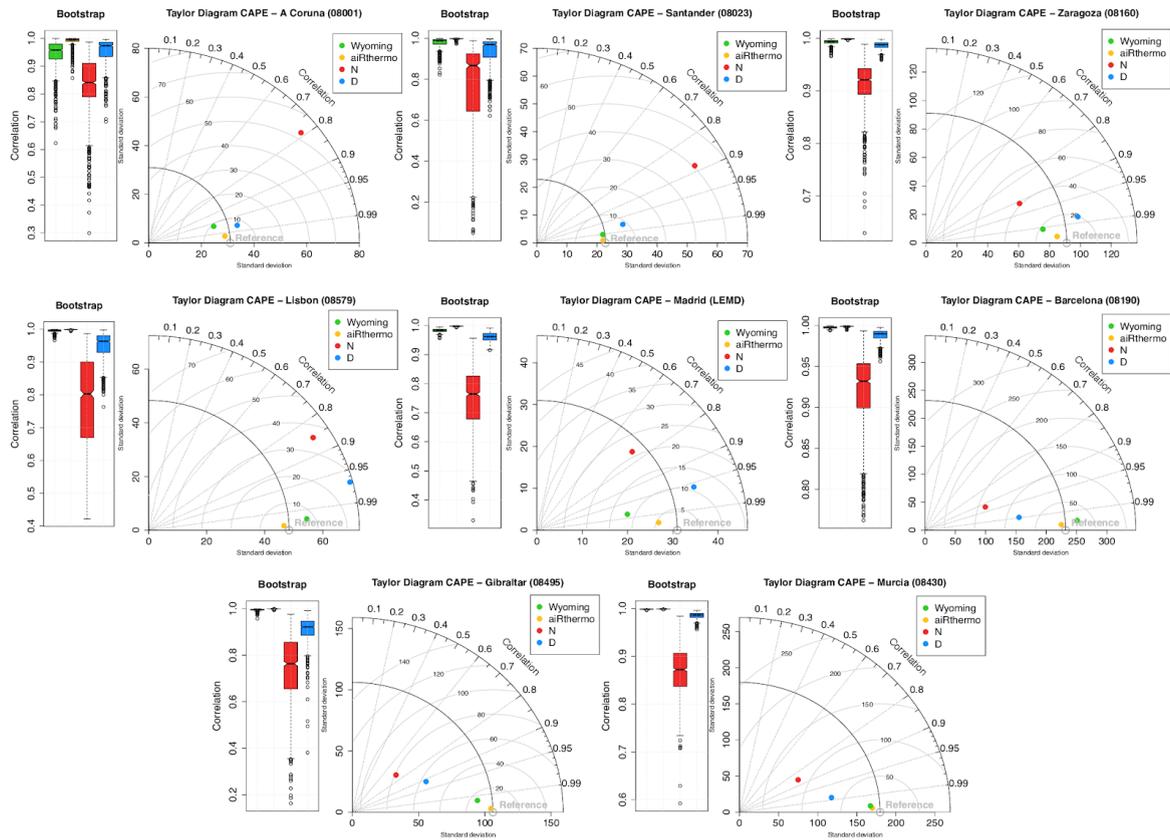


Fig. 2. Taylor diagrams of CAPE if the values from IGRA are taken as reference.

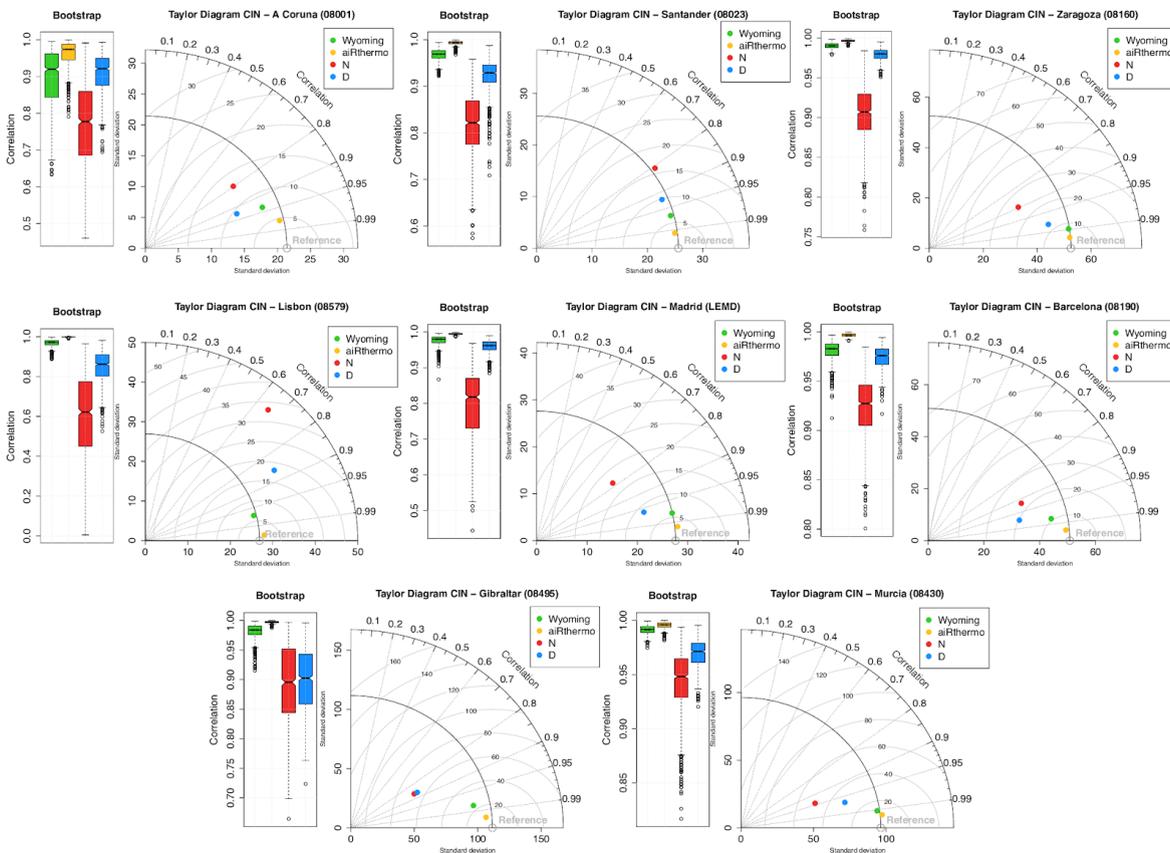


Fig.3. Taylor diagrams of CIN if the values from IGRA are taken as reference.

Comment made by: Anonymous Referee #1

Received and published: 28 April 2020

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

General comments:

This manuscript is very well-written and presents the results in a straight-forward manner. It is clear that assimilating various retrievals improves model representation of the three instability variables described. I only have a couple comments outlined below regarding the methodology details.

>> Thank you for these supportive words.

Comments:

Section 2.3.2: This section is very brief on the details of the bootstrapping technique. Perhaps include more details and some references?

>> The bootstrap operates by constructing the artificial data batches using sampling with replacement from the original data. Conceptually, the sampling process is equivalent to writing each of the N data values on separate slips of paper and putting all N slips of paper in a hat. To construct one bootstrap sample, N slips of paper are drawn from the hat and their data values recorded, but each slip is put back in the hat and mixed (this is the meaning of “with replacement”) before the next slip is drawn. This process is repeated a large number of times yielding, for example, to 1000 samples of size N that are slightly perturbed versions of the original data set. The 95% confidence intervals for the different statistical indicators can then be derived from their P975 and P025 observed percentiles of their distribution as obtained from the 1000 perturbed series. More mathematical details can be found in Wilks (2011), Efron & Gong (1983) and Downton & Katz (1993).

>> Coming back to our manuscript, the bootstrap technique was applied to the temporal analysis of each index. In our case, the original time series used in the Taylor diagrams consist of 60 values, each of them for the corresponding month along the period 2010-2014 (12 months x 5 years). For the bootstrap, we created 1000 perturbed time series taking into account different samples of the data. 67% of the new time series (2/3 of the length of the original time series - 40 values in our case) is made from the original data, and the remaining 33% (1/3 - 20 values) is chosen from those values already taken from the original data. For each correlation calculated, the same samples are taken from the observed and model data.

>> In order to clarify how the bootstrap is performed in our analysis, some extra lines will be added to section 2.3.2 (Analysis) of the paper. This new lines will sum up the information presented above, and they will include the citations.

>> Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). Academic press.

>> Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37(1), 36-48.
DOI: 10.1080/00031305.1983.10483087

>> Downton, M. W., & Katz, R. W. (1993). A test for inhomogeneous variance in time-averaged temperature data. *Journal of climate*, 6(12), 2448-2464.
DOI: 10.1175/1520-0442(1993)006<2448:ATFIVI>2.0.CO;2

What is exactly being verified? Model analyses after assimilation cycling or forecasts? This is not abundantly clear. If these are forecasts being verified, how do the statistics vary with lead time?

>> The structure of segments used in each experiment is presented in Figure 1. The N experiment is generated running 6-hour long segments that are restarted from the restart file produced at the end of previous segment (top panel of Figure 1). This is similar to a continuous WRF run where the boundary conditions (in our case, from ERA-Interim) are provided to the model every 6-hours after the initialization of the model the 1st of January, 2009.

>> For the experiment including the data assimilation, the structure is a little bit more complex. In these case, 12-hour long segments starting at every analysis time (00, 06, 12 and 18 UTC) are used (bottom panel of Figure 1). The analysis are generated from the outputs of the model at a 6-hour forecast step from the previous segment as first guess in a 3DVAR data assimilation scheme. The data assimilation is performed using the observations in PREPBUFR format obtained from the NCEP ADP Global Upper Air and Surface Weather Observations (ds337:0) dataset generated by NOAA. Only those observations included in a 2-hour time-window centered at the analysis times were included.

>> In both cases, the outputs are saved every 3 hours, which means that analysis (00, 06, 12 and 18 UTC) and 3-hour forecasts (at 03,09, 15 and 21 UTC) are included in our results. These recording frequency is highlighted with magenta ellipses in Figure 1. These are the data that are verified in the manuscript.

>> The new version of the manuscript will include an expanded explanation about how both simulations were created in order to make it clearer to the readers.

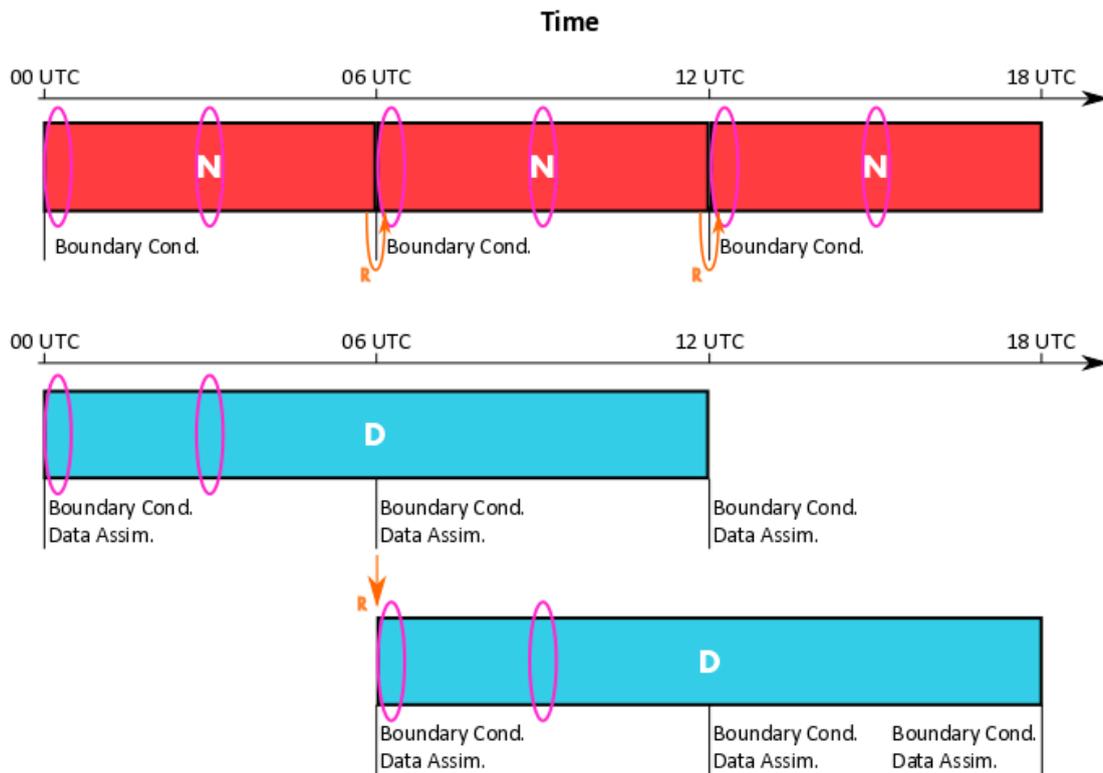


Figure 1: Diagram showing the structure of the segments used for running N (top, in red) and D (bottom, in blue) experiments with the WRF model. The outputs recorded are highlighted with magenta ellipses, and the restart files used to run the segments are shown in orange.

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 increase 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, how 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 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 of 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 findings 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 resolution 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-Rojá, 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.

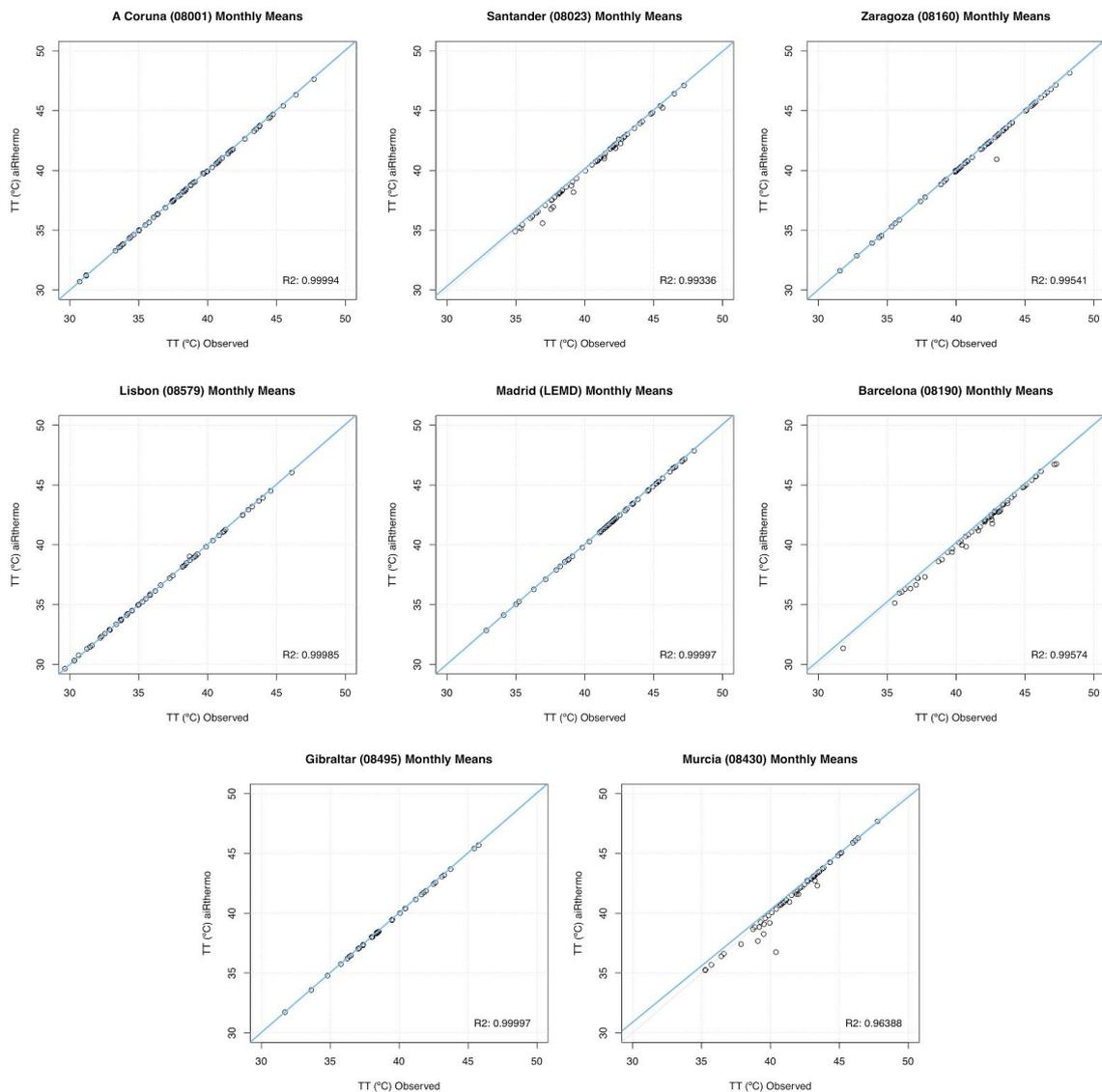


Figure 1: Scatterplots for the monthly values of the observed values of TT (directly taken from the University of Wyoming) against the monthly values of TT calculated with airRthermo. The R-squared is presented in the bottom right corner of each plot.

>> 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).

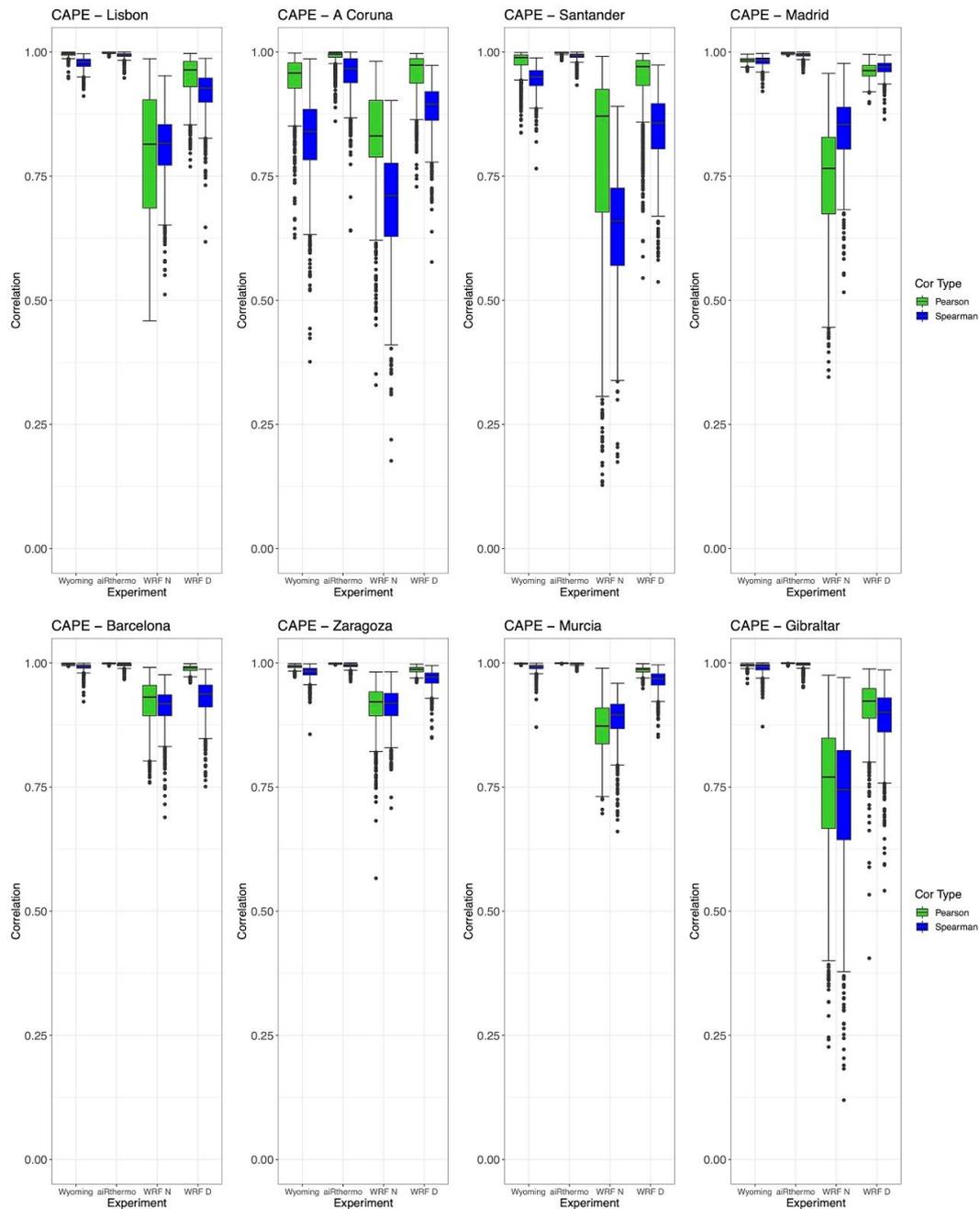


Figure 2: Box and Whiskers for the correlations obtained during the bootstrap for CAPE and calculated by different methods: Pearson in green, and Spearman in blue. IGRA stations are taken as Reference.

>> 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.

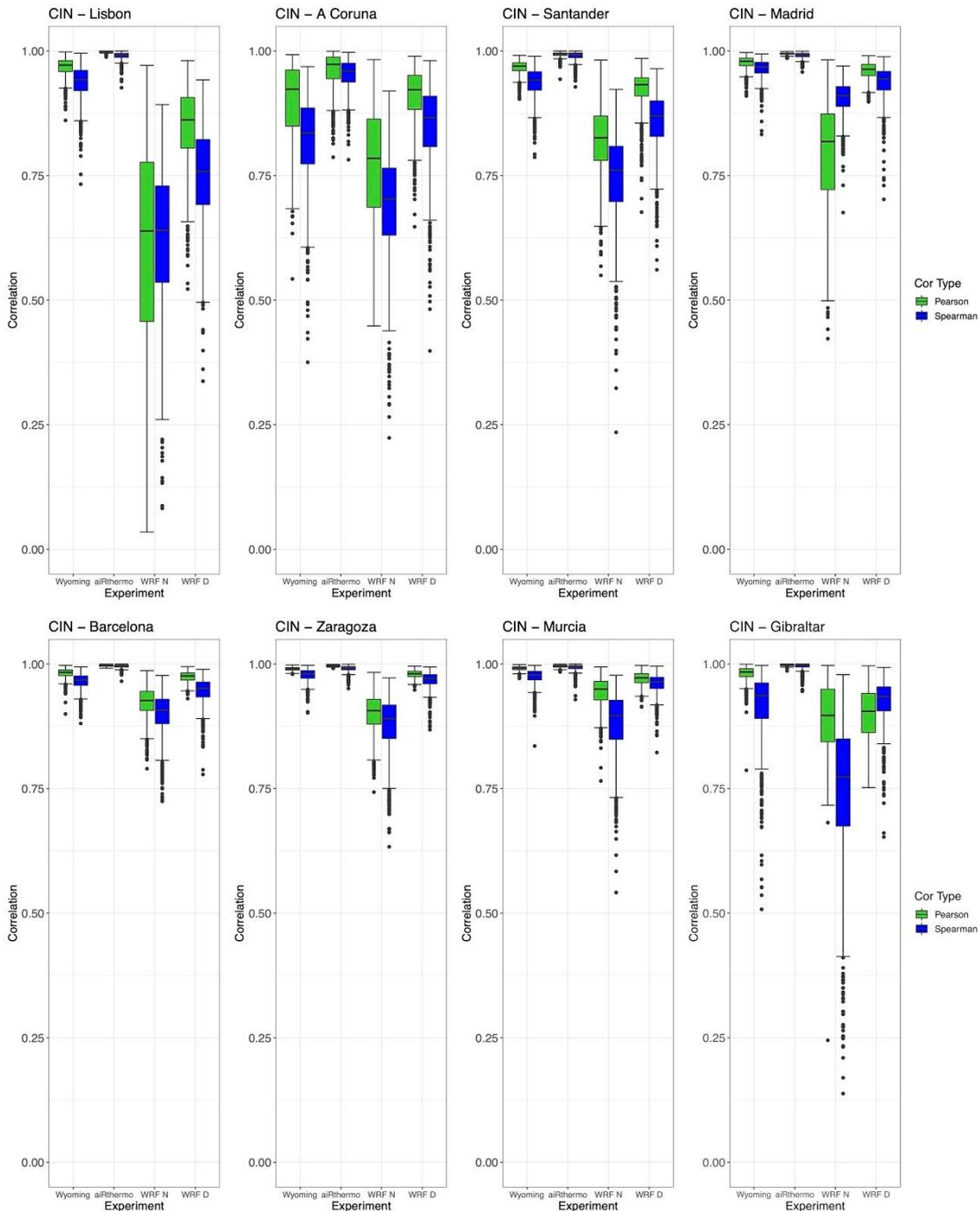


Figure 3: Same as Figure 2 but for CIN.

>> 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 rmse, 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 great 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 regimes 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.

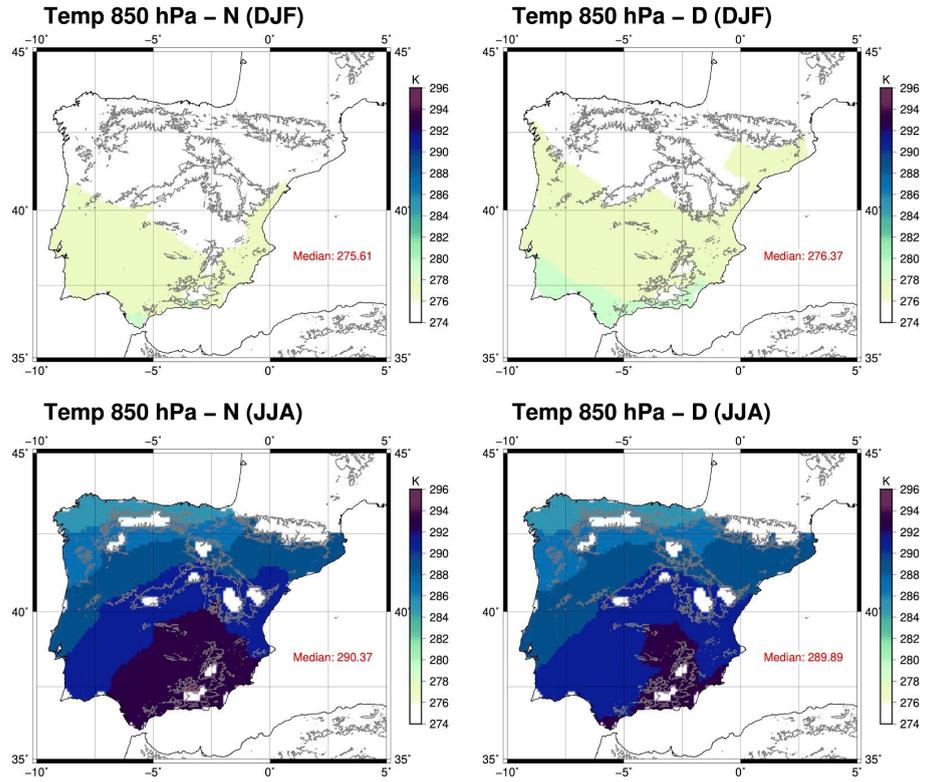


Figure 4: Spatial distribution of mean Temperature at 850 hPa for period 2010-2014 over the IP as computed from N (first column) and D (second column) for winter and summer. The median value (K) is in the bottom right corner of the plots.

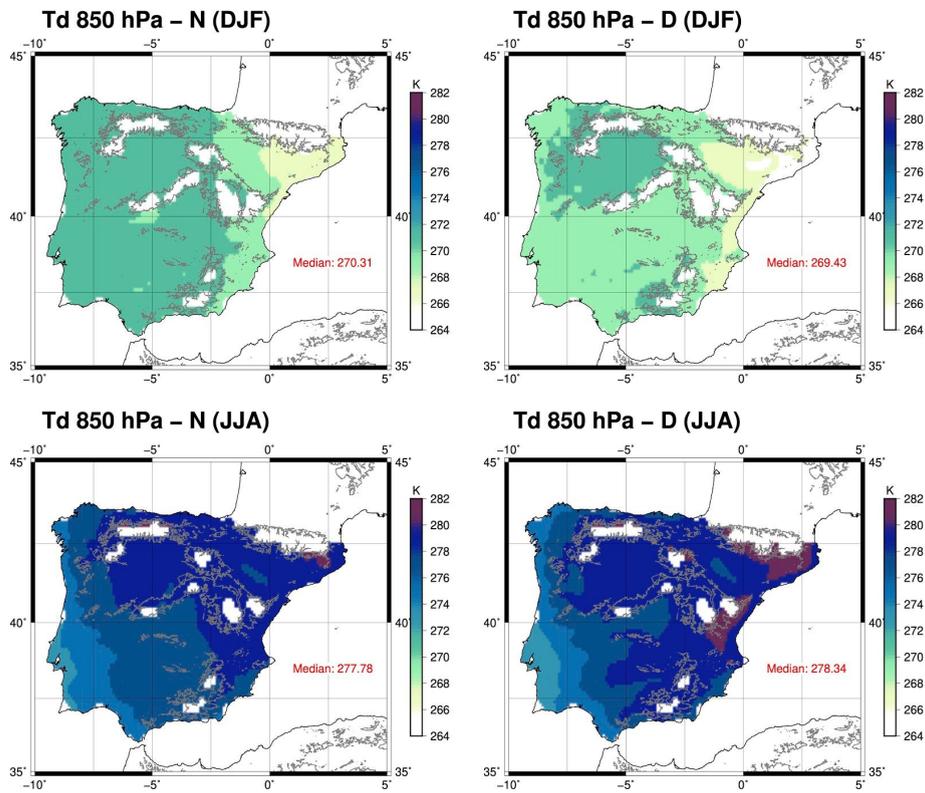


Figure 5: Same as Figure 4 but for dew point temperature at 850 hPa.

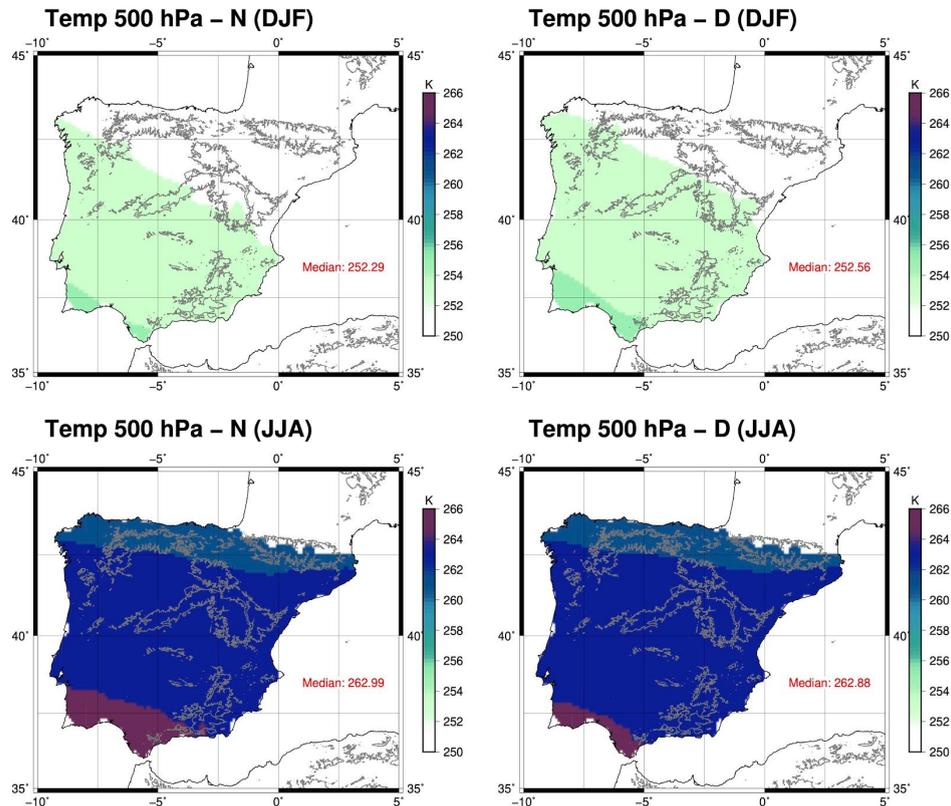


Figure 6: Same as Figures 4 and 5 but for temperature at 500 hPa.

>> 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 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.

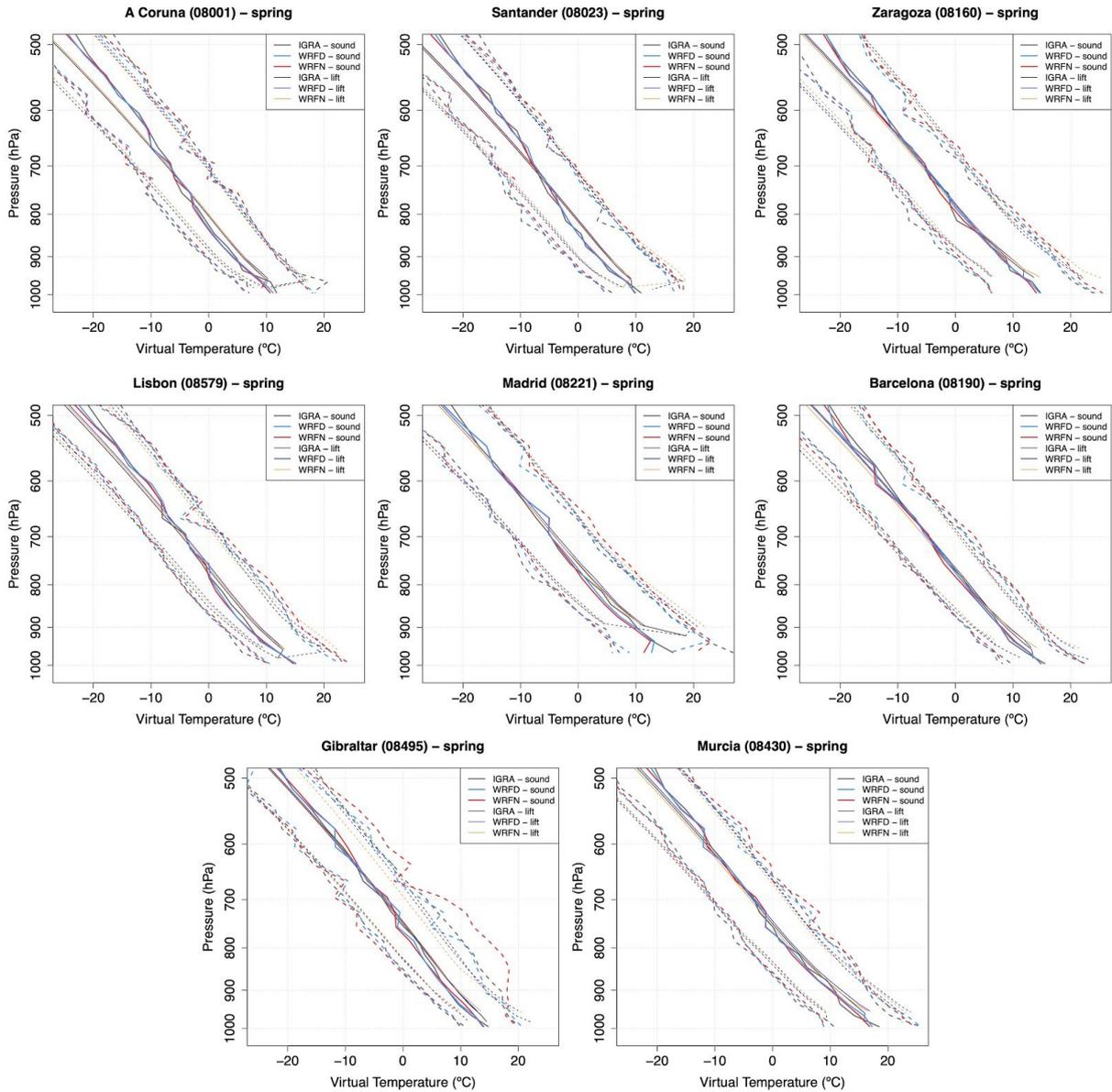


Figure 7: Vertical profiles of virtual temperature for the sounding levels and for the lifted parcel during spring. The dashed lines represent the 5 and 95 percentiles, and the solid lines the median.

>> 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.

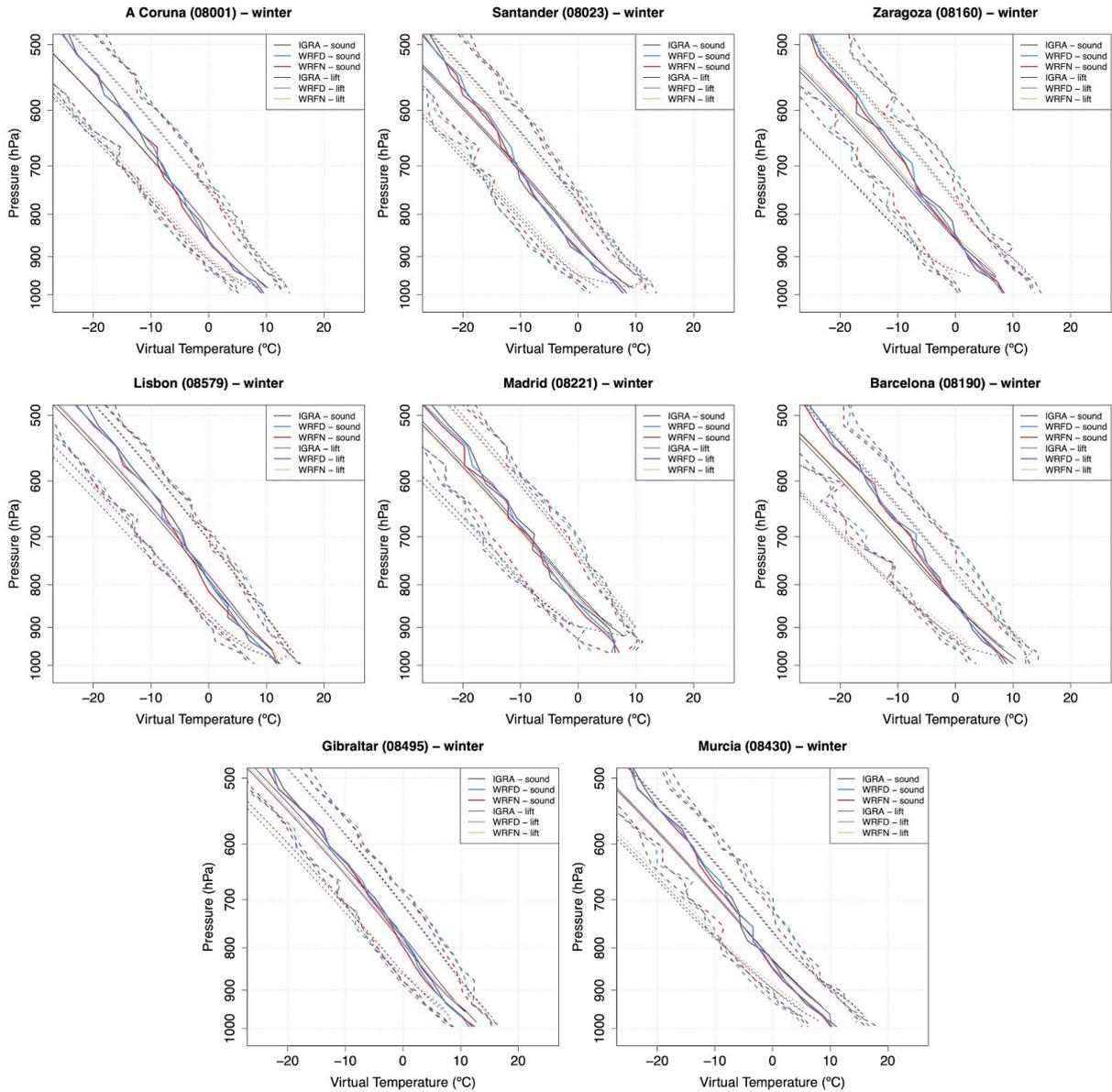


Figure 8: Same as Figure 7 but for winter.

>> 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.

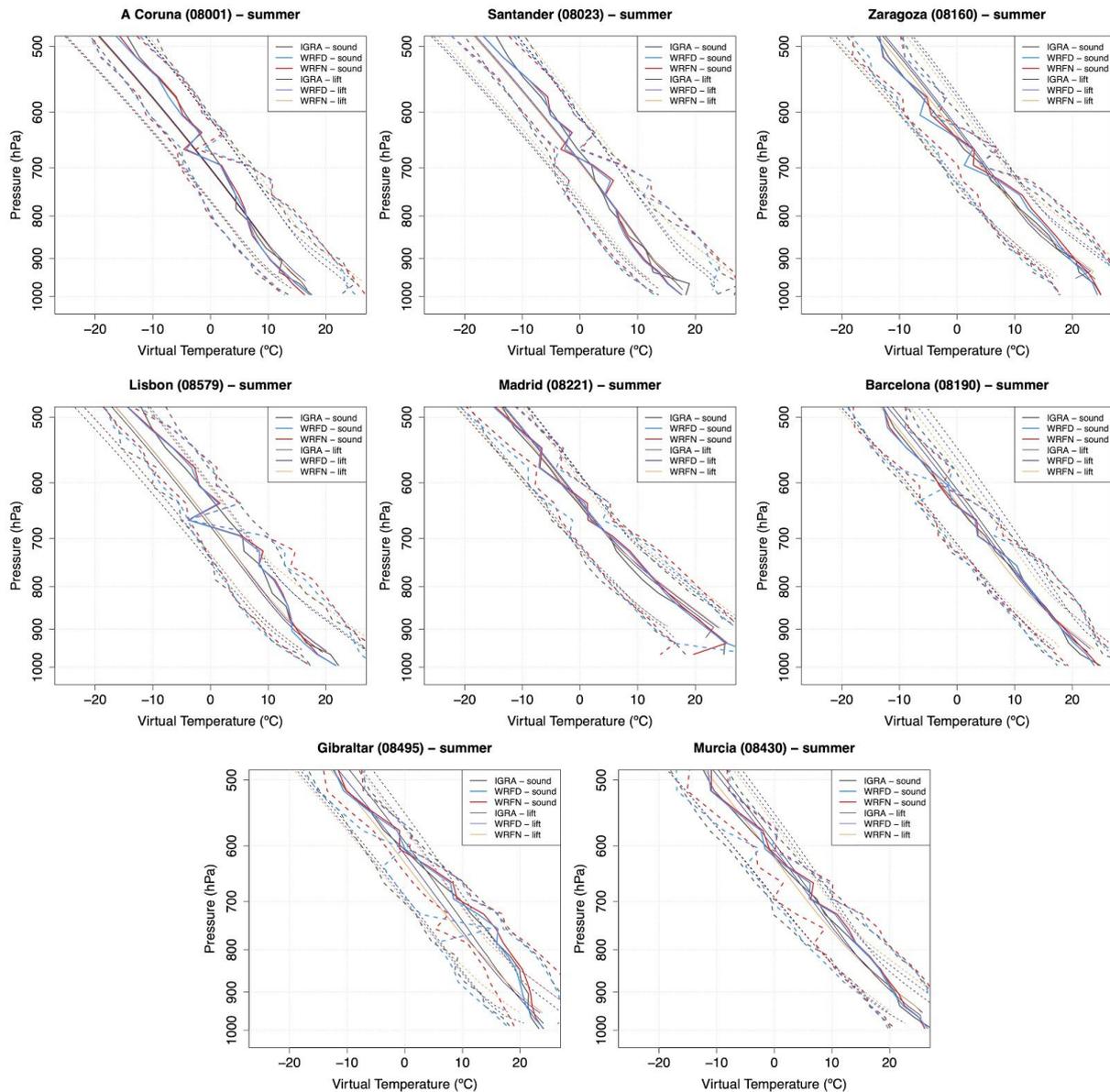


Figure 9: Same as Figures 7 and 8 but for summer.

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 reliable 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).

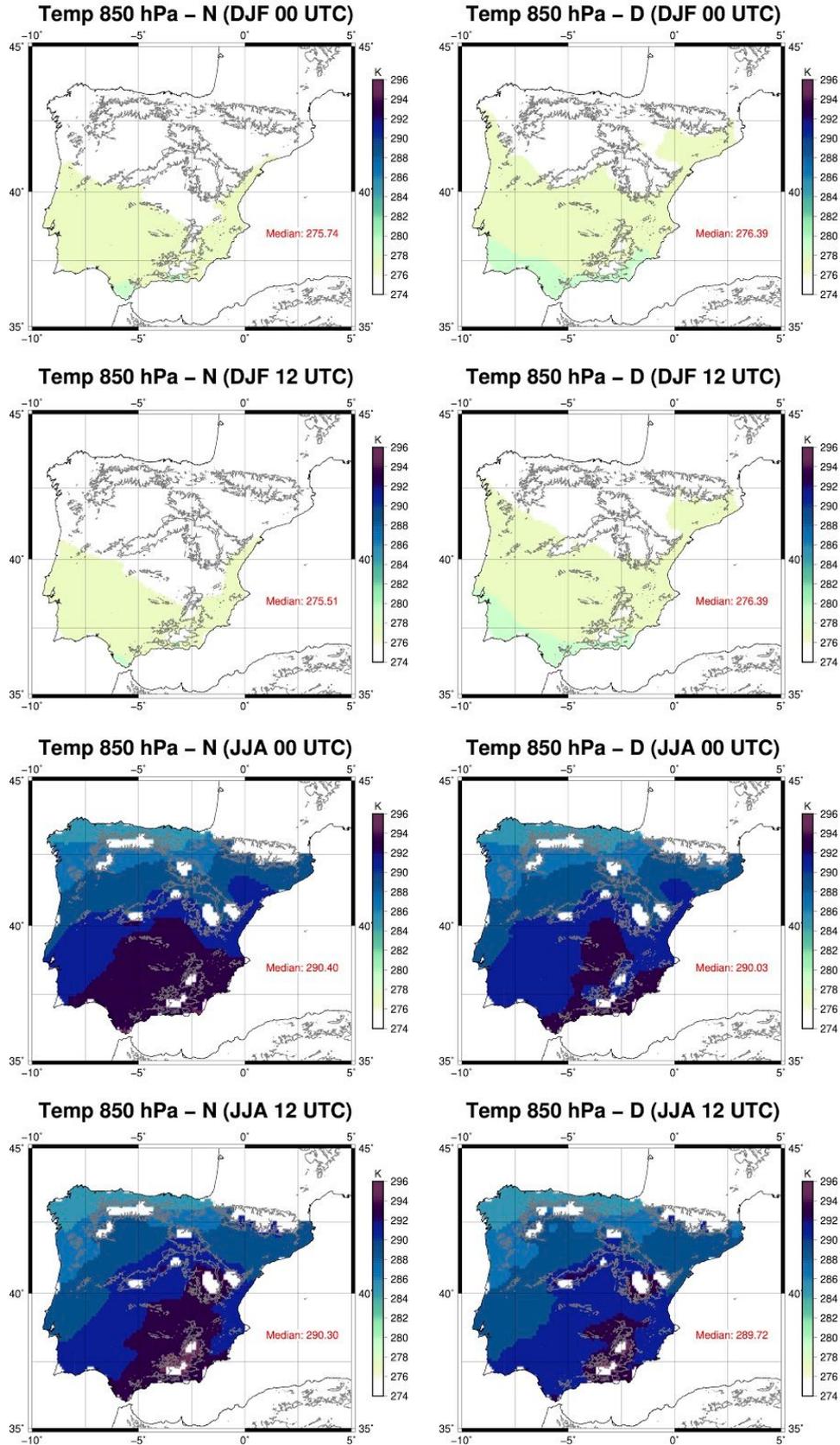


Figure 10: Spatial distribution of mean Temperature at 850 hPa for period 2010-2014 over the IP as computed from N (first column) and D (second column) for winter and summer at 00 and 12 UTC. The median value (K) is in the bottom right corner of the plots.

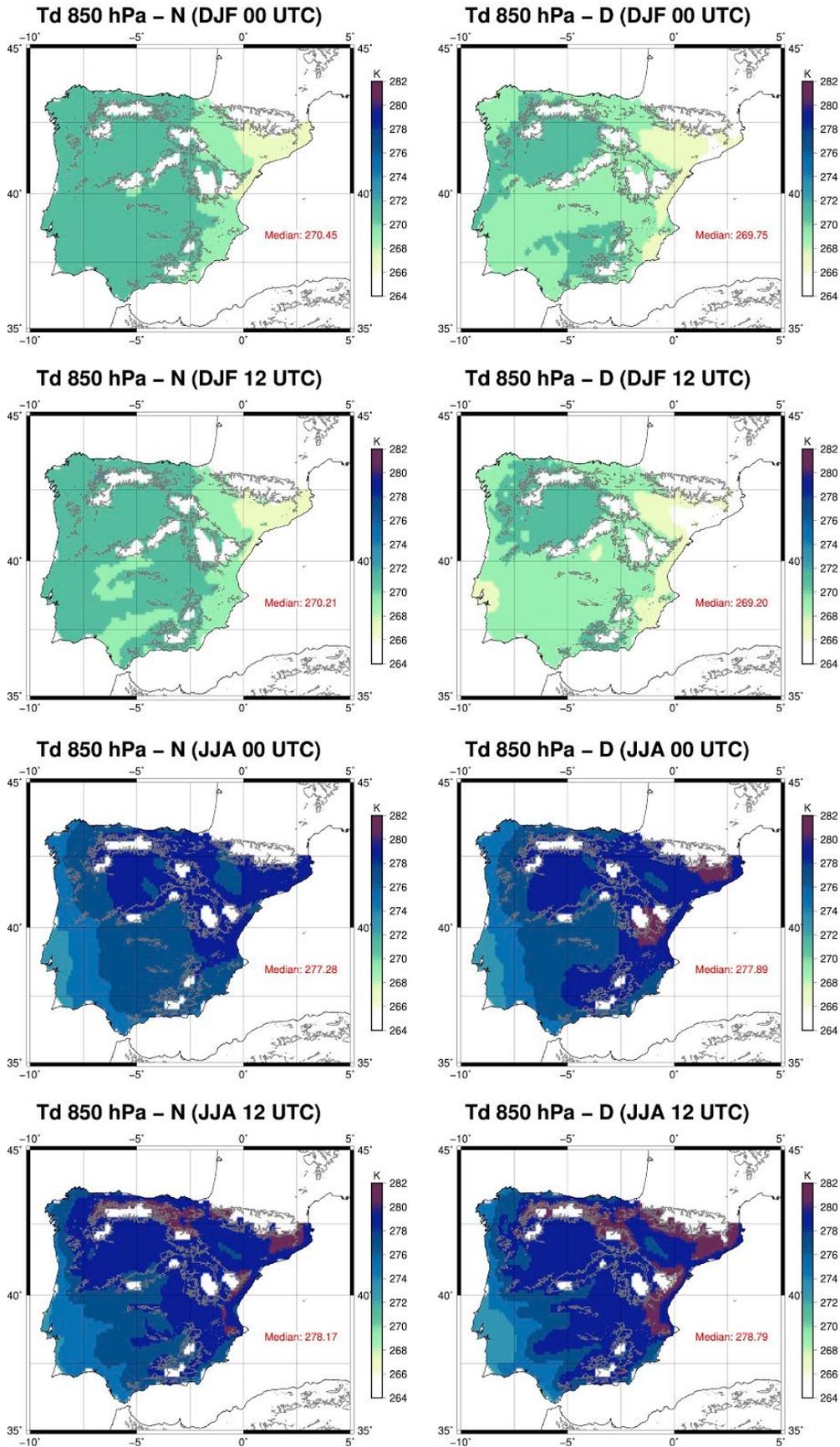


Figure 11: Same as Figure 10 but for dew point temperature at 850 hPa.

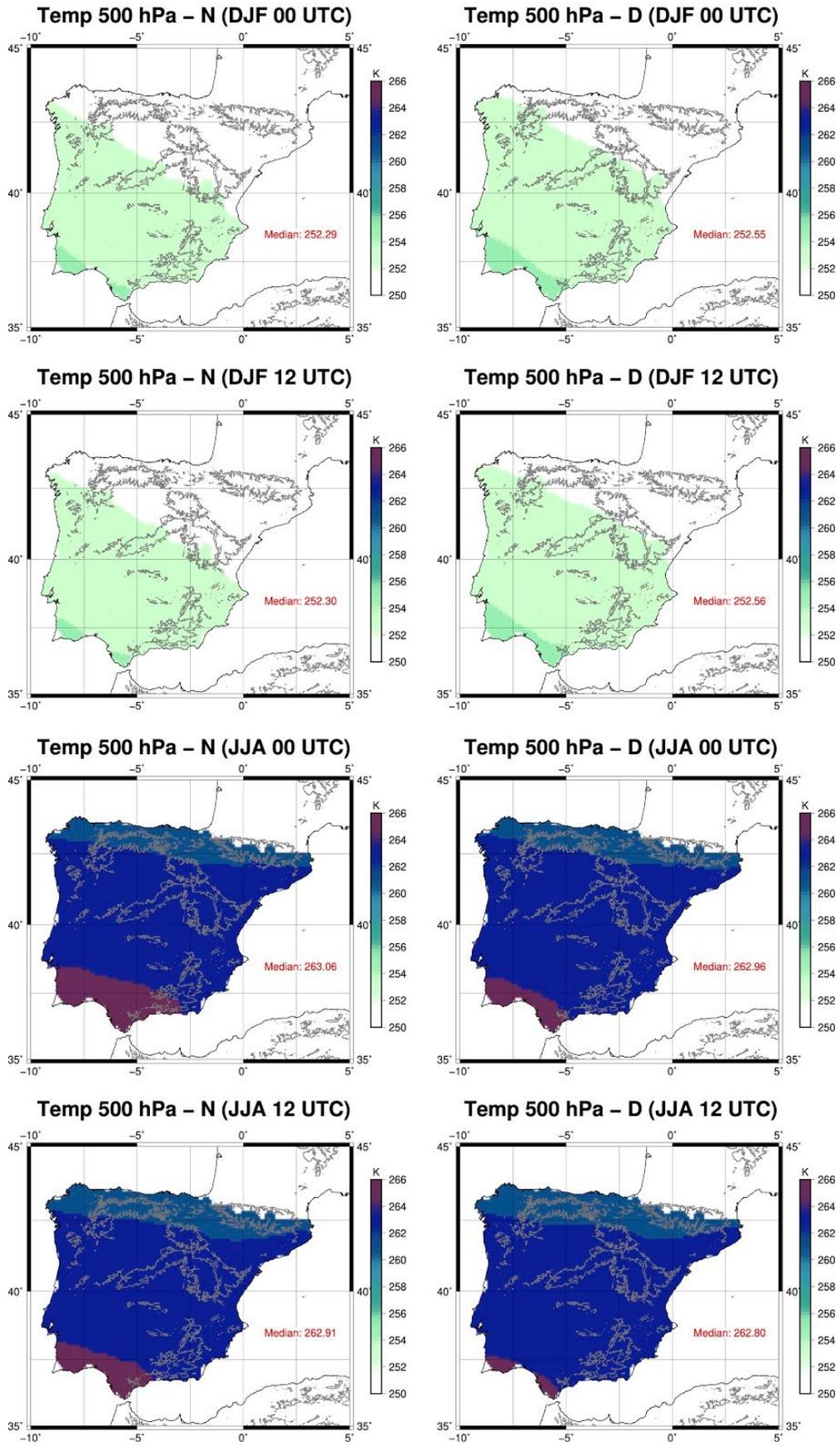


Figure 12: Same as Figures 10 and 11 but for temperature at 500 hPa.

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.

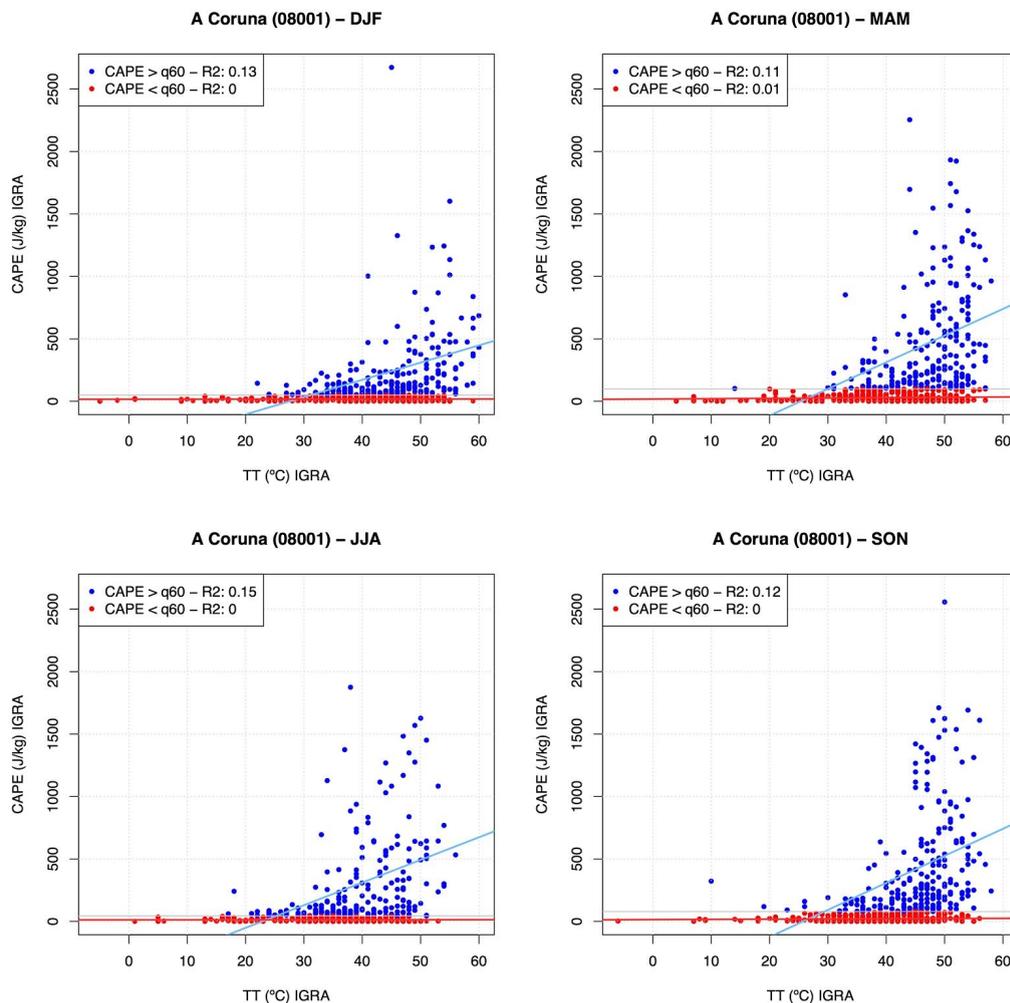


Figure 13: Scatterplots for the values of CAPE and TT as included in IGRA for A Coruna. The values of CAPE over the 60th percentile are in blue, and the values below that value are in red. The value of the 60th percentile is marked with a grey line, and the linear models are also included with the corresponding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

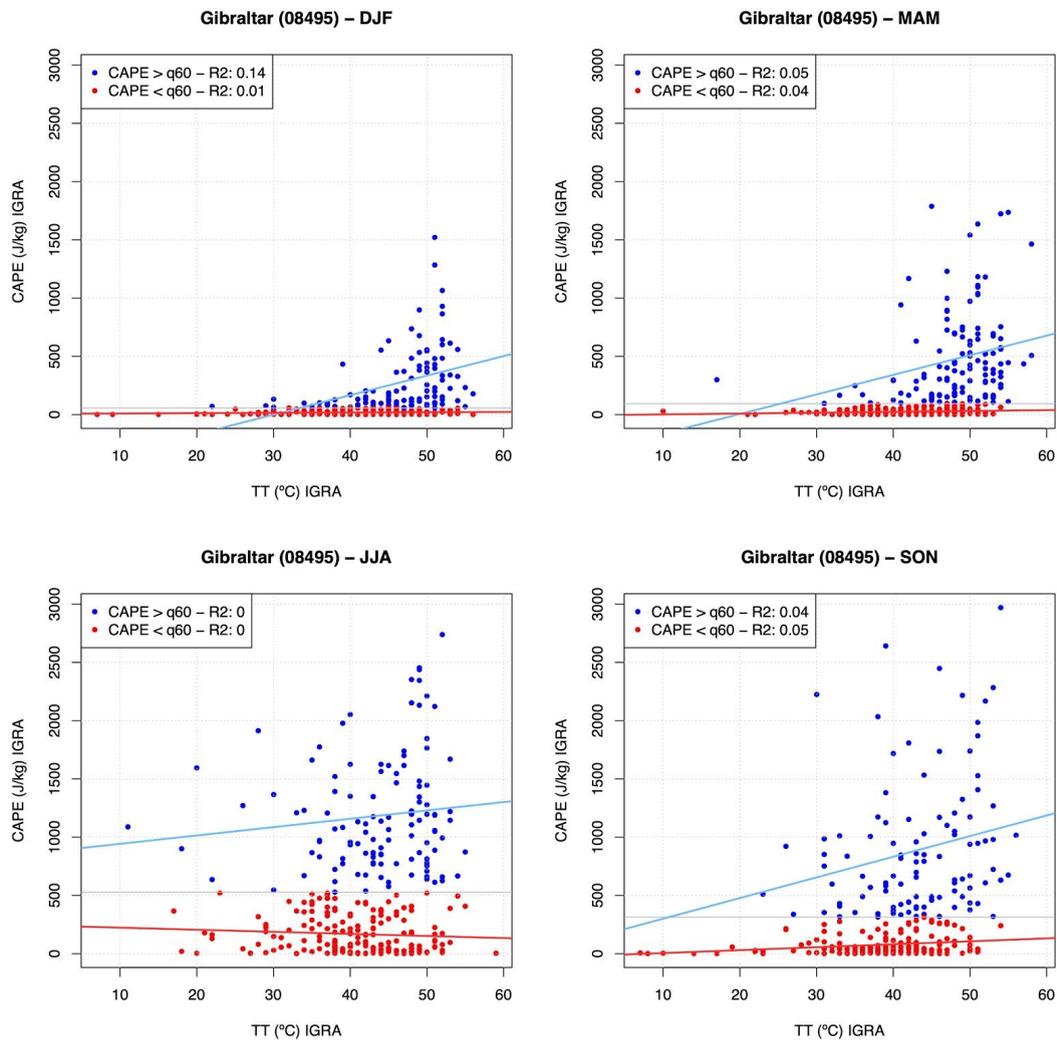


Figure 14: Same as Figure 13, but for Gibraltar.

>> 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 were 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. L123: 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.](#)

Changes in the simulation of instability indices over the Iberian Peninsula due to the use of 3DVAR data assimilation

Santos J. González-Rojí^{1,2}, Sheila Carreno-Madinabeitia^{3,4}, Jon Sáenz^{5,6}, and Gabriel Ibarra-Berastegi^{7,6}

¹Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland.

²Climate and Environmental Physics, University of Bern, Bern, Switzerland.

³Department of Applied Mathematics, Statistics and O.R., University of the Basque Country (UPV/EHU), Vitoria-Gasteiz, Spain.

⁴TECNALIA, Basque Research and Technology Alliance (BRTA), Parque Tecnológico de Álava, Vitoria-Gasteiz, Spain.

⁵Department of Applied Physics II, University of the Basque Country (UPV/EHU), Leioa, Spain.

⁶Plentzia Itsas Estazioa, PIE, University of the Basque Country (UPV/EHU), Plentzia, Spain.

⁷Department of NE and Fluid Mechanics, University of the Basque Country (UPV/EHU), Bilbao Engineering School, Bilbao, Spain.

Correspondence: Santos J. González-Rojí (santos.gonzalez@climate.unibe.ch)

Abstract. The ability of two downscaling experiments to correctly simulate ~~the instability conditions that can trigger thunderstorms~~ thermodynamic conditions over the Iberian Peninsula (IP) is compared in this paper. To do so, three instability indices are evaluated: TT index, CAPE and CIN. The WRF model is used for the simulations. The N experiment is driven by ERA-Interim's initial and boundary conditions; The D experiment has the same configuration as N, but the 3DVAR data assimilation step is additionally run at 00, 06, 12 and 18 UTC. Eight radiosondes are available over the IP, and the values for these indices ~~calculated from the University of Wyoming directly retrieved from the Integrated Global Radiosonde Archive (IGRA)~~ were chosen as reference in the validation of both simulations. Additionally, ~~measured variables at different pressure levels~~ vertical temperature and moisture profiles from the radiosondes provided by the University of Wyoming were used to calculate the three instability indices by our own methodology using the R package aiRthermo. According to the validation, the correlation, ~~SD and RMSE~~ Standard Deviation (SD) and Root Mean Squared Error (RMSE) obtained by the experiment D for all the indices in most of the stations are better than those for N. The different ~~methodologies~~ methods produce small discrepancies between the values for TT, but these are larger for CAPE and CIN due to the dependency of these indices on the initial conditions assumed for the calculation of ~~an air parcel's vertical evolution~~ a lifted air parcel. Similar results arise from the seasonal analysis concerning both WRF experiments: N tends to overestimate or underestimate (depending on the index) the variability of the reference values of the indices, but D is able to capture it in most of the seasons. The heterogeneity of the indices is highlighted in the mean maps over the ~~Iberian Peninsula~~ IP. According to those from D, the ~~ingredients for the development of convective precipitation during winter~~ unstable air masses are found along the entire Atlantic coast during winter, but in summer they are located particularly in over the Mediterranean coast. The ~~chances of developing thunderstorms~~ values of CAPE in those areas at 12 UTC is are much higher than at 00 UTC; The convective inhibition is more extended towards inland at 00 UTC in those areas, ~~which prevents storms from developing~~. However, high values are observed near ~~Murcia~~ the southeastern corner of the IP (near Murcia) also at 12 UTC.

1 Introduction

Precipitation is one of the most important variables involved in the water balance, and its variability determines the water resources ~~in of~~ the planet. ~~It~~ Following the definitions of regional models, precipitation can be separated in two categories: large-scale and convective precipitation, associated ~~with the frontal systems, and convective precipitation, triggered by convective instability of atmospheric layers. The latter for example with frontal systems and unstable atmospheric layers respectively.~~ In general, convective precipitation is usually associated with extreme events due to their high intensity and short duration. However, the simulation of these events is a well-known problem in the modelling community (Sillmann et al., 2013) due to restrictions in the resolution ~~and the~~, poor representation of complex topography ~~in the~~, forecast errors or deficiencies in the microphysics schemes in the numerical models. In order to avoid ~~this problem~~ these problems, as previously done in the literature (Viceto et al., 2017), this paper focuses on the evaluation of the atmospheric conditions favourable for the development of convective extreme precipitation rather than the validation of the simulation of extreme events.

The evaluation of the atmospheric conditions is typically based on the calculation of some instability indices such as Convective Available Potential Energy (CAPE) (Moncrieff, 1981), Convective INhibition (CIN) (Moncrieff, 1981), Lifted Index (LI) (Galway, 1956), K-Index (George, 1960), Total Totals index (TT) (Miller, 1975) or Showalter Index (S) (Showalter, 1953). All of these indices are commonly used in the literature for this kind of studies (Ye et al., 1998; DeRubertis, 2006; Viceto et al., 2017). CAPE and CIN are based on the adiabatic lifting of a parcel, while most of the others are based on differences in the values of several variables at different pressure levels. The deep convection and thunderstorms are triggered by three ingredients: high levels of moisture in the planetary boundary layer (PBL), potential instability and forced lifting (Johns and Doswell, 1992; McNulty, 1995; Holley et al., 2014; Gascón et al., 2015). CAPE and CIN provide information about the first two ingredients (Holley et al., 2014), and both can provide information about the genesis and intensity of the ~~convective precipitation~~ atmospheric convection (Riemann-Campe et al., 2009). However, previous studies (Angus et al., 1988; López et al., 2001) suggest that CAPE should not be used alone ~~and it~~, but should be combined with other indices. The final ingredient (lifting) is provided by the orography (Doswell et al., 1998; Siedlecki, 2009), the convergence of horizontal moisture fluxes (McNulty, 1995) or the breezes in coastal regions (van Delden, 2001). Thus, the high spatial and temporal resolution is important for ~~these studies~~ this kind of studies focusing on atmospheric convection, and that is why regional simulations are needed (Siedlecki, 2009).

The probability of occurrence of precipitation extreme events is not the same along the day, and previous studies ~~suggest~~ support that the maximum convection takes place in the afternoon and evening (~~Siedlecki, 2009~~) (Siedlecki, 2009; Virts et al., 2013; Piper and . According to van Delden (2001), the preferred time in most of ~~western~~ Western Europe is between 18 and 24 UTC, with the exception of the island of Corsica where the sea breeze triggers convection usually between 6-12 UTC. In open sea areas, the lightning activity peaks in the morning (Enno et al., 2020), associated to thunderstorms triggered by land breezes at night

(Virts et al., 2013). A regional study focusing over the UK (Holley et al., 2014) suggests that the reduction overnight of CAPE is over 500 J/Kg.

On the global scale, CAPE follows the spatial pattern of specific humidity and surface air temperature, which means that it increases from pole to Equator (Riemann-Campe et al., 2009). The minimums are obtained in arid regions and over areas with cold water up-welling. Focusing on Europe, convective precipitation is developed with storms develop for lower values than the U.S. (Kaltenböck et al., 2009), and several studies tried to determine the most active regions. Romero et al. (2007) found that the most unstable region is located along a zonal belt over the south-central Europe, particularly in over the west Mediterranean sea and the surrounding areas. This agrees with Brooks et al. (2003), who found that the favourable environment for thunderstorms is developed in the southern Europe, and that the highest number of days in such a regime is located in central Europe, are located over the Iberian Peninsula (hereafter, IP), south of the Alps, and southern Ukraine Northern Balkans. However, van Delden (2001) found that the southwestern France and the Basque Country seem to be a preferred region for the formation of severe storms that drift towards the northeast. More recent studies based on lightning data (Enno et al., 2020) and regional climate models using higher resolution (Rädler et al., 2018; Mohr et al., 2015) highlighted the same areas with favourable environments for thunderstorms in Europe. Particularly, over northern Italy (Po Valley), east of the Adriatic Sea (Albania, Bosnia and Serbia) and in the northeastern IP and Southern France (near the Gulf of Lyon).

Over the Iberian Peninsula (hereafter, IP) Over the IP, the seasonality of precipitation is determined by its heterogeneous topography and the different sources of moisture affecting the region due to seasonal variations of the global atmospheric circulation and contrasting climatic regions (influenced by the strong topography). Northern and western IP are mainly affected by stratiform precipitation during winter, while eastern and southern IP receive great amounts of precipitation during autumn due to convective activity (Rodríguez-Puebla et al., 1998; Esteban-Parra et al., 1998; Romero et al., 1999; Iturrioz et al., 2007). Maximum precipitation amounts over central IP are measured in early spring (Tullot, 2000).

Previous studies about instability indices over the IP (Viceto et al., 2017) suggest that CAPE presents high spatio-temporal shows a high spatiotemporal variability: the values in winter and spring over land are small due to the reduced surface temperature, and the differences between Atlantic and Mediterranean regions are remarkable during summer. According to Siedlecki (2009), they the mean values range from below 50 J/kg in the north to between 100 and 200 J/kg in the Mediterranean coast (some events can even reach 1000 J/Kg). As Romero et al. (2007), Viceto et al. (2017) also stated that CAPE is low during autumn in the Atlantic and continental regions, but high in the areas surrounding the Mediterranean sea. This seasonality was also observed for other indices such as K-index or TT, which show maximum values during summer (Siedlecki, 2009). Observations proved that annual precipitation over eastern stations is mostly accumulated during autumn, as a result of the cumulative warming of the Mediterranean sea due to summer insolation (Romero et al., 2007; Iturrioz et al., 2007), and later entrance entry of very hot and humid air into the IP while cold air is present at higher levels (Dai, 1999; Eshel and Farrell, 2001; Correoso et al., 2006). Additionally, September and October are the months with highest frequency of waterspouts and tornadoes near the Balearic Islands (Gayà et al., 2001). Over the northwestern northwestern IP, the mean CAPE when a hailstorm occurs hailstorms occur is 360 J/kg, while for a thunderstorm thunderstorms it is only 259 J/kg (López et al., 2001). The values are

similar to those observed in Europe, but the dispersion of the dispersion of these values is really high (almost 350 J/kg over the whole sample)(Alexander and Young, 1992; Lucas et al., 1994).

90 ~~Recent studies (Brooks, 2013; Rädler et al., 2019) suggest that due~~, which is similar to that found in previous studies (Alexander and Young, 1992; Lucas et al., 1994). The values are similar to those observed in other regions of Europe, but lower than those values obtained in studies based on synoptic or lightning data for severe hailstorms (around 500 J/kg) (Kunz, 2007; Púčik et al., 2015; Taszarek et al., 2017). Due to global warming, the conditions necessary for the development of extreme precipitation events will be enhanced (Brooks, 2013; Rädler et al., 2019). The frequency and intensity of climate
95 extremes will be magnified (Diftenbaugh et al., 2013). ~~Marsh et al. (2009) projected higher~~, projecting larger values of CAPE in the Mediterranean during summer and autumn ~~, which is in line with the projections from Viceto et al. (2017) for the IP and its surroundings. Thus, the probability of observing lower values of CAPE is expected to decrease in the future.~~ (Marsh et al., 2009; Viceto et al., 2017).

The main objective of this paper is to evaluate the performance of two simulations created by using the Weather and Research
100 Forecasting (WRF) model (Skamarock et al., 2008) (including or not the extra 3DVAR data assimilation step) at reproducing the atmospheric conditions that can trigger convective precipitation over the IP. ~~To do so, the~~ if the third ingredient (e.g. lifting) is fulfilled. We are not restricting our analysis only to extreme events, and the entire period from 2010-2014 will be considered. For the evaluation, the comparison of pseudo-soundings extracted from the model against real observations will be carried out. Additionally, the seasonal patterns of different instability indices will be studied. The novelty of this study lies in the inclusion
105 of data assimilation in the downscaling experiment used for the analysis of some instability indices.

This paper is organised as follows: The details of the configuration of the WRF model used in both experiments are presented in section 2, along with a brief outline of the methodologies used in the study. The main results are presented in section 3, while they are compared against previous studies presented in the introduction. Finally, we conclude with some remarks about our research in section 4.

110 2 Data and Methodology

2.1 WRF model configuration

Two experiments were carried out using version 3.6.1 of the WRF model for the period 2010-2014. In both simulations, ERA-Interim provides the initial and boundary conditions (Dee et al., 2011). 6-hourly data at 0.75 degrees were downloaded from the Meteorological Archival and Retrieval System (MARS) repository at European Centre for Medium-Range Weather Forecasts
115 (ECMWF). ~~Analyses of temperature, relative humidity, both wind components and geopotential at 20 pressure levels (5, and they present a 0.75 degrees resolution and 10, 20 vertical levels, 30, 50, 70, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 850, 900, 925, 950, 1000 hPa) were used to feed WRF.~~ Both simulations were started on the 1st of January, 2009 from a cold start. Following similar methodologies to previous studies (Argüeso et al., 2011; Zheng et al., 2017), the entire year 2009 was selected as the spin-up for the land surface model included in WRF, and consequently, it was omitted in the study presented
120 here.

One of the experiments (hereafter, N) was nested inside ERA-Interim as usual in numerical downscaling ~~exercises~~experiments, which means that the model is driven by the boundary conditions after its initialization. It is generated running 6-hour long segments that are restarted from the restart file produced at the end of previous segment, which is similar to a continuous WRF run where the boundary conditions are provided to the model every 6-hours after the initialization of the model. The other experiment (D) presents the same ~~configuration~~parameterizations as N, but with the additional 3DVAR data assimilation step (Barker et al., 2004, 2012) that is run every 6 hours (at 00, 06, 12 and 18 UTC). In this ~~stepcase~~, 12-hour long segments starting at every analysis time (00, 06, 12 and 18 UTC) are used. The analyses are generated from the outputs of the model at a 6-hour forecast step from the previous segment as first guess in a 3DVAR data assimilation scheme. In both experiments, the outputs are saved every 3 hours, which means that analyses (00, 06, 12 and 18 UTC) and 3-hour forecasts (at 03, 09, 15 and 21 UTC) are included in our results. In the data assimilation step, quality controlled temperature, moisture, pressure and wind observations in PREPBUFR format from the NCEP ADP Global Upper Air and Surface Weather Observations dataset (referenced as *ds337.0* at NCAR's Research Data Archive) were included. Only those observations included in a time-window of two-hours centered in the analysis times were assimilated.

As Figure 1 shows, the domain focuses over the IP, but it also includes parts of Europe, Africa and the Atlantic ocean. As stated by previous studies (Jones et al., 1995; Rummukainen, 2010), the set-up of the domain used in this study prevents border-effects affecting our results as mesoscale systems can develop freely. The spatial resolution of both experiments is 15 km, and ~~they use with~~ 51 vertical levels. ~~The recording frequency of the outputs is 3 hours in both experiments~~up to 20 hPa in eta (η) coordinates.

Apart ~~form~~from the ERA-Interim data, sea surface temperature (SST) of the model was updated on a daily basis using the high-resolution dataset *NOAA OI SST v2* (Reynolds et al., 2007). Additionally, the following parameterizations for the physics of the model were included in both WRF simulations: five-class microphysics scheme (WSM5) (Hong et al., 2004), MYNN2 planetary boundary layer scheme (Nakanishi and Niino, 2006), Tiedtke cumulus convection scheme (Tiedtke, 1989; Zhang et al., 2011), RRTMG scheme for both long and shortwave radiation (Iacono et al., 2008), and NOAH land surface model (Tewari et al., 2004).

The background error covariance matrices were created before running the simulation with 3DVAR data assimilation. To do so, the CV5 method included in WRFDA (Parrish and Derber, 1992) was used. A separate simulation initialized at 00 and 12 UTC and spanning 13 months (from January 2007 to February 2008) was necessary for the calculation of these matrices. Independent matrices were created for each month, and each of them was calculated taking into account a 90 days period centered on each month.

Both simulations were already presented and validated in previous studies by the authors. Precipitable water, precipitation and evaporation over the IP were validated against station measurements and gridded datasets in González-Rojí et al. (2018), and the outputs produced by D were always superior to N and the driving reanalysis ERA-Interim (for the latter, at least comparable for some variables). The closure of the water balance was also better for D. Additionally, the precipitation from D exhibited similar capabilities to the one downscaled with statistical methods (González-Rojí et al., 2019). Furthermore, the wind field from D showed also improvements compared to ERA-Interim, and consequently, that data were used for the

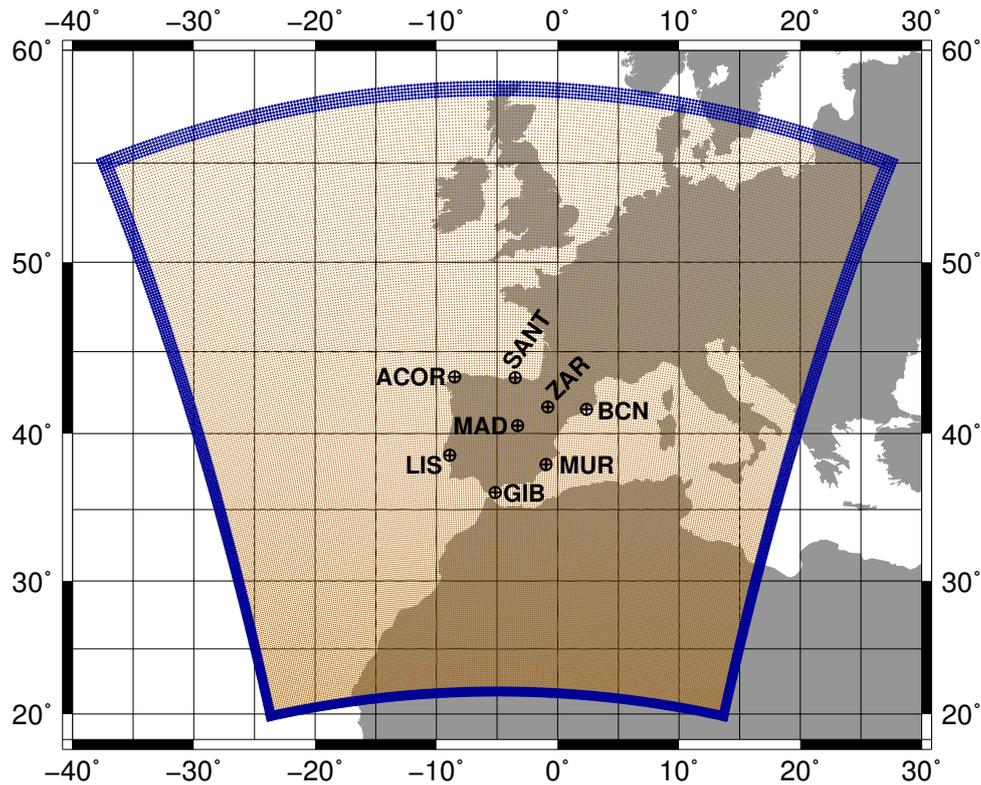


Figure 1. The domain used in both WRF simulations is presented with darkorange dots, while the darkblue region highlights the relaxation zone. The location of all the radiosondes available over the IP is also presented with quartered circles.

calculation of the offshore wind energy potential in the west Mediterranean (Ulazia et al., 2017). Afterwards, that study was extended to every coast of the IP (Ulazia et al., 2019). The moisture recycling over the IP was also evaluated in González-Rojf et al. (2020), highlighting its importance in the Mediterranean coast during spring and summer.

2.2 Radiosonde data

160 Atmospheric radiosonde data were downloaded from the server of the University of Wyoming (freely accessible at <http://weather.uwyo.edu/upperair/sounding.html>). Even if the University of Wyoming does not apply any quality control to the data, this dataset was already used in previous studies by the authors and none of the values were taken as erroneous. Additionally, data from the Integrated Global Radiosonde Archive (IGRA) created by NOAA was also included in this study. This dataset is available online after several quality control procedures.

165 Only eight radiosondes are available over the IP: A ~~Coruña~~ Coruña (ACOR), Santander (SANT), Zaragoza (ZAR), Barcelona (BCN), Madrid (MAD), Lisbon (LIS), Gibraltar (GIB) and Murcia (MUR). The location of each station is presented in Figure 1. Measurements are carried out every day at midday and midnight (00 and 12 UTC, 02 and 14 LT), with the exception of Lisbon

where they ~~only measure it are only available~~ at 12 UTC (13 LT). Additionally, the amount of data available for Gibraltar is extremely scarce since August 2012.

170 ~~Height, pressure, temperature, potential temperature~~ Temperature and mixing ratio were retrieved at all the available pressure levels at each location. Moreover, the values of the TT, CAPE and CIN ~~indices~~ as calculated by the University of Wyoming were also ~~downloaded. These values were assumed as the reference in our analysis.~~

~~Additionally, retrieved. Additionally, vertical profiles of~~ pressure, temperature and mixing ratio ~~at different pressure levels obtained download~~ from the University of Wyoming were also used to calculate ~~the~~ TT, CAPE and CIN ~~indices~~ following our own methodology using the aiRthermo R package (further details can be found in the next subsection). The comparison between the original values of the indices retrieved ~~from the University of Wyoming~~ and our results can give us information about whether their discrepancies are only due to differences in the calculation procedure. The values of pressure, temperature and moisture at each level from IGRA were used to calculate TT, CAPE and CIN following our own methodology employing aiRthermo. The values computed from the IGRA dataset were assumed as the reference in our analysis.

180 It must be said that all the radiosondes presented here were assimilated during the 3DVAR data assimilation step in WRF. ~~As stated in González-Rojí et al. (2018), the analysis increments are stronger~~ The effect of the data assimilation step was measured by the analysis increments (analysis minus background) in González-Rojí et al. (2018). The effect of the data assimilation is more intense at 12 UTC, particularly in southeastern IP and both Guadalquivir and Ebro basins UTC compared to the other times, and particularly for summer (see their Figure 13). ~~This pattern is consistent along the seasons, but its intensity varies seasonally. Strong increments are observed during summer, but not during winter. Thus, the assimilation is useful to correct~~ The spatial analysis of these values highlighted that the effect of data assimilation is not homogeneous over the IP, and it concentrates mainly in the southeastern IP and both Guadalquivir and Ebro basins. This is related to the well-known cold bias observed during summer over the IP in the IP in summer in WRF simulations (Fernández et al., 2007; Argüeso et al., 2011; Jerez et al., 2012) ~~and the effect of the assimilation is not restricted only to the station location, as the data assimilation is able~~ to make it much smaller.

2.3 Methodology

2.3.1 Calculation of instability indices

For both simulations, the nearest grid point to the real latitude and longitude of each radiosonde was determined, and the corresponding pseudo-sounding (pressure, ~~potential~~ temperature and mixing ratio) at 00 and 12 UTC was obtained at ~~model~~ WRF's original ~~eta levels. η levels.~~ We did not consider the average value of several grid points (e.g., an array of 3×3 grid points) as we would be taking into account an area of 2025 km^2 ($45 \text{ km} \times 45 \text{ km}$), and that would be too much for a comparison against radiosonde data. Additionally, 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 (Xu et al., 2015). That means that our spatial resolution is suitable for the direct comparison of the nearest points against radiosonde data, as we do not neglect the horizontal drift of the radiosoundings.

200 This procedure was tested for reanalysis (Lee, 2002) and model data (Molina et al., 2020), and it showed that these pseudo-soundings are able to reproduce reasonably well the atmospheric conditions measured by the soundings. However, as highlighted by Holley et al. (2014), this procedure takes into account a stationary column at a fixed time, which can influence the comparison to real radiosonde data ~~because they are made by balloons that~~ as these measurements are not instantaneous and not in a straight vertical line (the balloons take many minutes to measure the profile of the atmosphere and ~~that deviate from a~~
205 straight vertical line they are deviated because of wind).

In order to calculate the instability indices TT, CAPE and CIN using the ~~pseudo-sounding~~ pseudo-soundings from the model, the R package *aiRthermo* was used (Sáenz et al., 2019). The most recent version was used (version 1.2.1), which is publicly available in the CRAN repository (<https://cran.r-project.org/package=aiRthermo>). Both CAPE and CIN are calculated by means of the vertical integrals using discrete slabs defined by the resolution of pressure in the soundings ~~-.The (using~~
210 all the available levels). 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 virtual temperature was used in every integral (Doswell and Rasmussen, 1994). ~~The energy on each slab is calculated analytically, and the values are accumulated until the end of the sounding, which means that the final value of CAPE and CIN are obtained. Further details about the functions used for the calculation of the vertical evolution of the air parcels can be found in Sáenz et al. (2019) and~~
215 also in the manual of the R package *aiRthermo* associated to that publication.

Additionally, in order to calculate CAPE and CIN in the most similar way to the University of Wyoming with the aim of reducing the differences between the values due to different calculation procedures (Siedlecki, 2009), the average of the lower vertical levels was set as the initial representative parcel (Craven et al., 2002; Letkewicz and Parker, 2010). As in Siedlecki (2009), the averaged values from the lowest 500 m were used in this study. ~~Additionally, Furthermore, in order to avoid that~~
220 the averaged initial parcel state is still too hot compared to the ambient conditions (in that case, CIN will never be computed as the parcel is already artificially buoyant), an isobaric precooling was applied ~~to the initial parcel state~~ if needed. To do that, the parcel was cooled along an isobar until it crossed the sounding so that it was not buoyant.

The TT index was calculated following the definition from Miller (1975). It is defined as:

$$TT = (T_{850} - T_{500}) + (D_{850} - T_{500}) \quad (1)$$

225 where T_{850} and T_{500} are the temperatures ~~in Celsius~~ at 850 and 500 hPa, and D_{850} is the dewpoint temperature ~~in Celsius~~ at 850 hPa. According to the ECMWF (Owens and Hewson, 2018), thunderstorms are likely when the values for this index are above 44 °C. However, other values can be found in the literature: 48.1°C for southern Germany (Kunz, 2007), 46.7°C for the Netherlands (Haklander and Van Delden, 2003) or 46°C for Switzerland (Huntrieser et al., 1997).

It can be seen that this index is not highly dependant on the initial conditions for its calculation as it only depends on
230 temperature at ~~different pressure levels. two discrete pressure levels, while~~ CAPE and CIN are very sensitive to the initial conditions used for the simulated ascent. TT avoids this problem, but the results can suffer from errors due to inversion layers (Siedlecki, 2009). It must be pointed out that the dew point temperature is needed for TT, and that it is highly important for the calculation of the Lifted Condensation Level (LCL) while calculating CAPE and CIN. In the case of the radiosonde data, the

235 indices are calculated using the measured dew point temperature at 850 hPa when is needed, while in our method, this variable
is calculated from the temperature and mixing ratio at that pressure level. This can trigger small differences in the results even
if the same original radiosonde data are used.

Further indices could be calculated from the pseudo-soundings obtained from the outputs of the model or real observations, but they were omitted because they can provide similar information to CAPE, CIN and TT. K-index is also based on temperature at different pressure levels, so it suffers from the same problems as TT. Additionally, previous studies reported a strong
240 correlation between CAPE and LI (Blanchard, 1998; López et al., 2001). In order to avoid these connections between indices, we restricted this study to TT, CAPE and CIN indices.

2.3.2 Analysis

Once TT, CAPE and CIN are calculated at the nearest grid points to radiosonde locations of both simulations (N and D), and also those using the original sounding data (labelled as *aiRthermo* ‘*aiRthermo*’ in the results), we obtain a time series with a 12-
245 hourly temporal resolution for each index. These values can be compared against the reference values of the indices retrieved directly from IGRA and the University of Wyoming (labelled as *Reference* ‘*Reference*’ and ‘*Wyoming*’ in the next figures). The comparison of *Reference vs aiRthermo* ‘*Wyoming*’ vs ‘*aiRthermo*’ aims to achieve an estimation of the error/differences due to the different ~~methodologies used~~ methods applied by both sources of results. This comparison was based on independent locations over the IP (separated by several kilometers), so a Taylor diagram was chosen as the best option to show Pearson’s
250 correlation (r), ~~root-mean-squared-error~~ Root Mean Squared Error (RMSE) and ~~standard-deviation~~ Standard Deviation (SD) of each experiment in the same plot. In order to determine which experiment is doing the best job at simulating the reference values of the indices, the procedure explained by Taylor (2001) was followed: the dots that lie nearest to the reference on the X axis represent variables that agree well with observations (high correlations and low RMSEs), and those lying near the highlighted arc will present comparable standard deviations to the observations.

255 Additionally, the bootstrap technique with resampling was applied to the results in order to represent an estimation of the sampling errors from each experiment. ~~In this case,~~ (Efron and Gong, 1983; Wilks, 2011). In our case, the original time series used in the Taylor diagrams consist of 60 values, each of them for the corresponding month along the period 2010-2014 (12 months \times 5 years). For the bootstrap, we created 1000 perturbed time series taking into account different samples of the data. 67 % of the new time series ~~were subsampled with replacement~~ (2/3 of the length of the original time series – 40 values
260 in our case) is made from the original data, and the remaining 33 % (1/3 – 20 values) is chosen from those values already taken from the original data. For each correlation calculated, the same samples are taken from all datasets and experiments. The variability of the ~~correlations of~~ Pearson’s correlations obtained with these synthetic time series were shown by means of ~~box and whiskers plots.~~ Box-Whiskers plots. To complete the statistical analysis of the bootstrap, the Spearman correlations were also added to these Box-Whiskers plots.

265 Then, the seasonal analysis of each index at each location was carried out. In this case, the variability of the results is showed by different ~~box and whiskers~~ Box-Whiskers plots. Each season was defined as follows: winter is defined from December to

February (DJF), spring from March to May (MAM), summer from June to August (JJA) and autumn from September to November (SON).

270 Finally, the calculation of each index was extended to every grid point included in a mask defined for the land points of the IP over model's domain. The spatial distribution of the mean values of each index at 00 and 12 UTC during winter and summer was calculated. These maps show the ~~regions over the IP were the convective precipitation can be developed and in which~~ spatial distribution of TT, CAPE and CIN over the land grid points in the IP which are more prevalent in each season.

3 Results

275 Taylor diagrams for the TT index calculated for each radiosonde of the IP are shown in Figure 2. The ~~box and whiskers~~ Box-Whiskers associated to the correlations (both Pearson's and Spearman's) obtained for each of the 1000 time-series created with the bootstrap technique are also included. According to the Taylor diagrams, as expected, the best experiment reproducing the reference values is Wyoming, followed by aiRthermo (the real measurements of temperature, mixing ratio and pressure from the sounding were used to calculate TT with our methodology), ~~followed by~~ D and later by N. ~~aiRthermo~~ Wyoming obtains the closest values to the observations in all the stations ~~with the exception of Murcia.~~ The results from aiRthermo are quite similar
280 to Wyoming, except for Murcia where D is ~~the best one reproducing the observed~~ better reproducing the reference data. The correlations are always above 0.99 for Wyoming, 0.98 for aiRthermo, 0.97 for D and 0.75 for N. The observed SD is really well simulated by Wyoming, aiRthermo and D, but N underestimates it in most of the stations as it is only able to reproduce the one in Santander and A Coruna. The RMSE is below 0.6 °C for aiRthermo, below 1 °C for D and below 2.5 °C for N.

285 The bootstrap analysis is consistent with the results obtained in the Taylor diagrams, and it shows that the Pearson's correlations are always above 0.99 for Wyoming and 0.98 for aiRthermo (again, with the exception of Murcia where they are above 0.9). The correlations are always above 0.95 for D. In the case of N, the spread of the values is much larger than for aiRthermo and D, and their ~~mean median~~ values are obtained between 0.8 and 0.9. If we change to Spearman's correlations, we can see that values are similar but with a small worsening of the values (particularly in Gibraltar, Murcia and Madrid).

290 Thus, as expected, we obtain the most similar results to those calculated from ~~the University of Wyoming~~ IGRA (Reference) with the values from Wyoming and those calculated with the real measurements from the soundings (that is, aiRthermo). However, we can still measure small differences between the values of the datasets due to the ~~different procedure in use of~~ measured dew point temperature in Wyoming, but the calculation of ~~the TT index~~ it from temperature and mixing ratio in IGRA and aiRthermo. These differences are remarkable in Murcia. Between both WRF experiments, it is clear that the experiment including the 3DVAR data assimilation is able to outperform the standard simulation only driven by the reanalysis data at the
295 boundaries of the domain. The differences between both WRF simulations are highlighted, particularly in those stations located in the Mediterranean coast (Barcelona, Murcia and Gibraltar) and in Lisbon.

In the case of CAPE, the validation results are presented in Figure 3. ~~Similar results to what we obtained for TT index is observed for CAPE: the~~ The best experiment reproducing the results is aiRthermo, followed by Wyoming, D and finally by N. The correlations are in all the stations above ~~0.96 in aiRthermo~~ 0.99 in aiRthermo and 0.95 for Wyoming, while for D they are

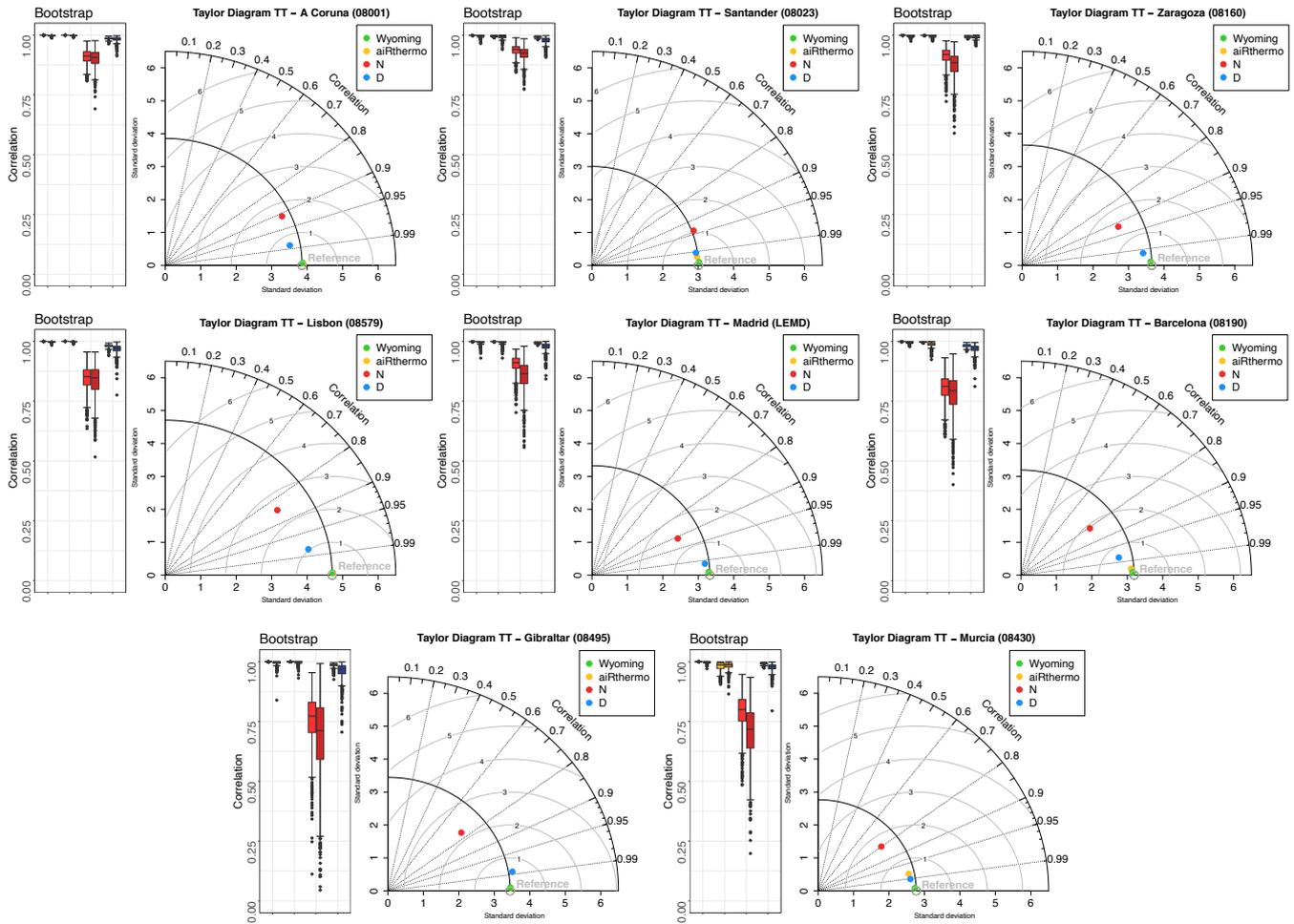


Figure 2. Taylor diagrams showing the r , RMSE and SD values for [Wyoming](#), [aiRthermo](#), [N](#) and [D](#) compared to TT values retrieved directly from [the University of Wyoming IGRA](#) (Reference). On the left side of each Taylor diagram, a [box-and-whisker Box-Whiskers](#) plot is added in order to show [the Pearson's and Spearman's](#) correlations between each experiment and the reference data ([lighter and darker colours, or first and second columns of the Box-Whiskers](#) respectively). The bootstrap technique with resampling was used to create 1000 synthetic time-series. [Wyoming](#), [aiRthermo](#), [N](#) and [D](#) are plotted in [green](#), orange, red and blue respectively.

300 above 0.9 and above [0.65-0.7](#) for [N](#). A similar behaviour is observed for SD and RMSE. The largest RMSEs are obtained in Barcelona and Murcia (both in the Mediterranean region), [where the RMSEs for N are 157.03 J/kg and 105.97 J/kg respectively \(98.32 J/kg for D and 30.32 J/kg for aiRthermo in Barcelona, 55.62 J/kg for D and 8.23 J/kg for aiRthermo in Murcia\).](#)

The bootstrap analysis shows that the highest [Pearson's](#) correlations are obtained by [aiRthermo](#) ([as expected](#)) and [Wyoming](#), but followed really closely by [D](#). As for TT index, [N](#) presents the worst performance and the largest spread. [The correlations can](#)
 305 [be even negative for some outliers in Santander \(up to -0.2\)](#) [If we consider, instead, the use of Spearman's correlations, we can](#)

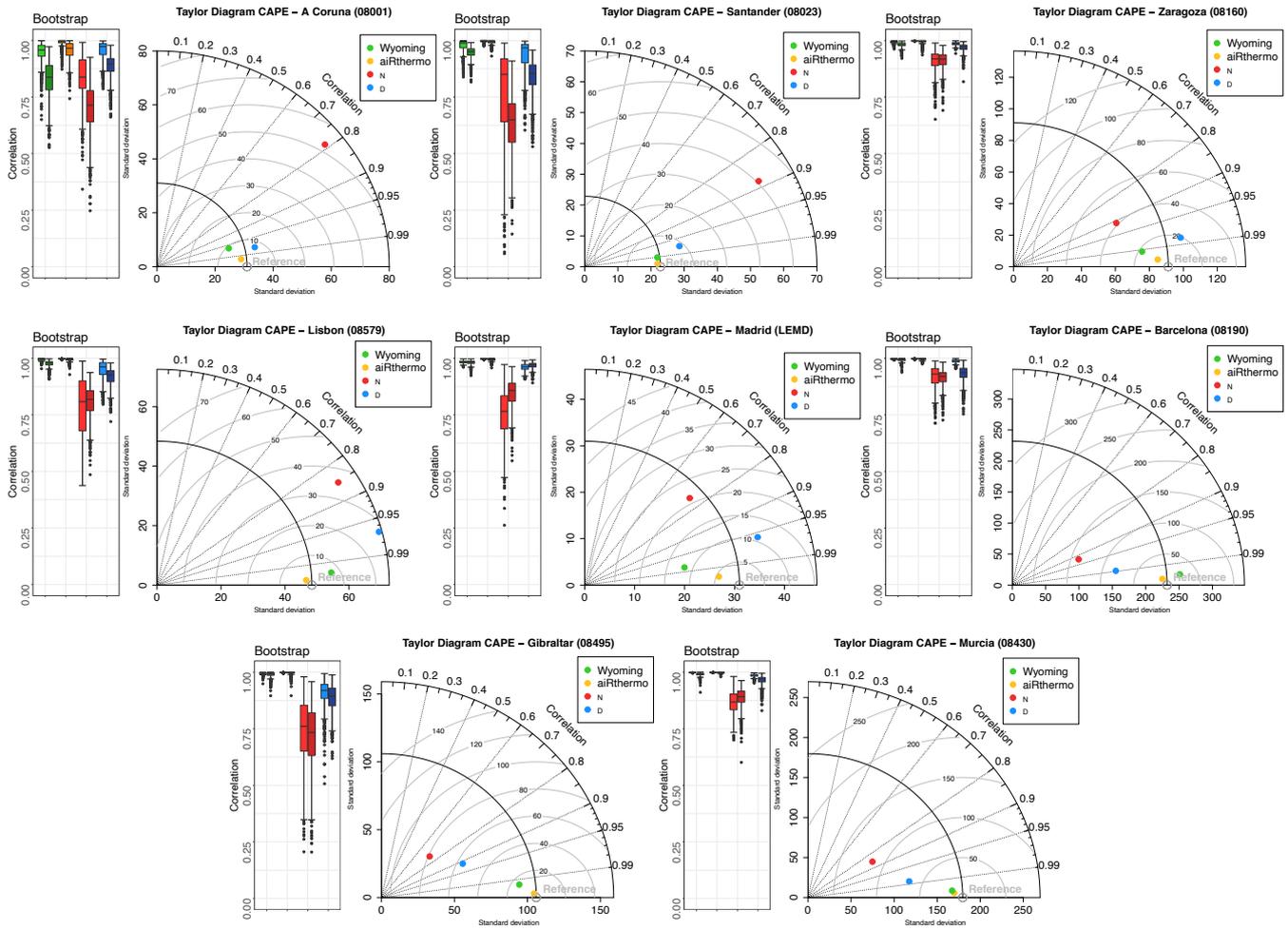


Figure 3. Same as Figure 2 but for CAPE.

[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.](#)

As stated before in section 2.3.1, the calculation of CAPE is ~~much~~ more difficult than TT index, and this is highlighted in the validation of these results. Even if the same data are used for the calculation of CAPE ([Reference](#)-[Wyoming](#) and aiRthermo used the same measurements), it is clear that small differences in the initial conditions can [trigger result in](#) serious discrepancies between both methods as stated by Siedlecki (2009). The largest RMSEs for aiRthermo can be found in Barcelona and Murcia ([30.32 and 8.23 J/kg respectively](#)). As for the TT index, two stations in the Mediterranean coast present the largest differences between [the](#) experiments. However, while the computation of TT from both WRF simulations produces standard deviations similar to the observed ones, the results for CAPE substantially overestimate the variance of Atlantic sites

315 (A ~~Coruña~~Coruña, Santander, Lisbon and Madrid) or overestimate it in the Mediterranean (Barcelona, Murcia and Gibraltar).
 Anyway, it can be seen that data assimilation improves the simulation of CAPE over the IP.

Finally, the validation of CIN is presented in Figure 4. As ~~expected for CAPE~~, the best results are obtained again by aiRthermo, followed by Wyoming, D and N (with the exception in Gibraltar where D and N are really similar). aiRthermo obtains in every station correlations above ~~0.950.97~~, followed by Wyoming, D and N (~~correlations above with correlations above 0.93~~, 0.85 and 0.65 respectively). ~~aiRthermo tends to slightly overestimate the SD, but both~~. Both WRF experiments overestimate or underestimate it depending on the station (particularly N in Lisbon, Madrid, Murcia and Zaragoza). The RMSE is always larger for N, and the values are remarkable in Murcia and Gibraltar for all the experiments (~~Gibraltar: aiRthermo 25.42 J/kg, D 52.67 J/kg, N 53.47 J/kg; Murcia: aiRthermo 10.71 J/kg, D 28.35 J/kg, N 48.24 J/kg~~).

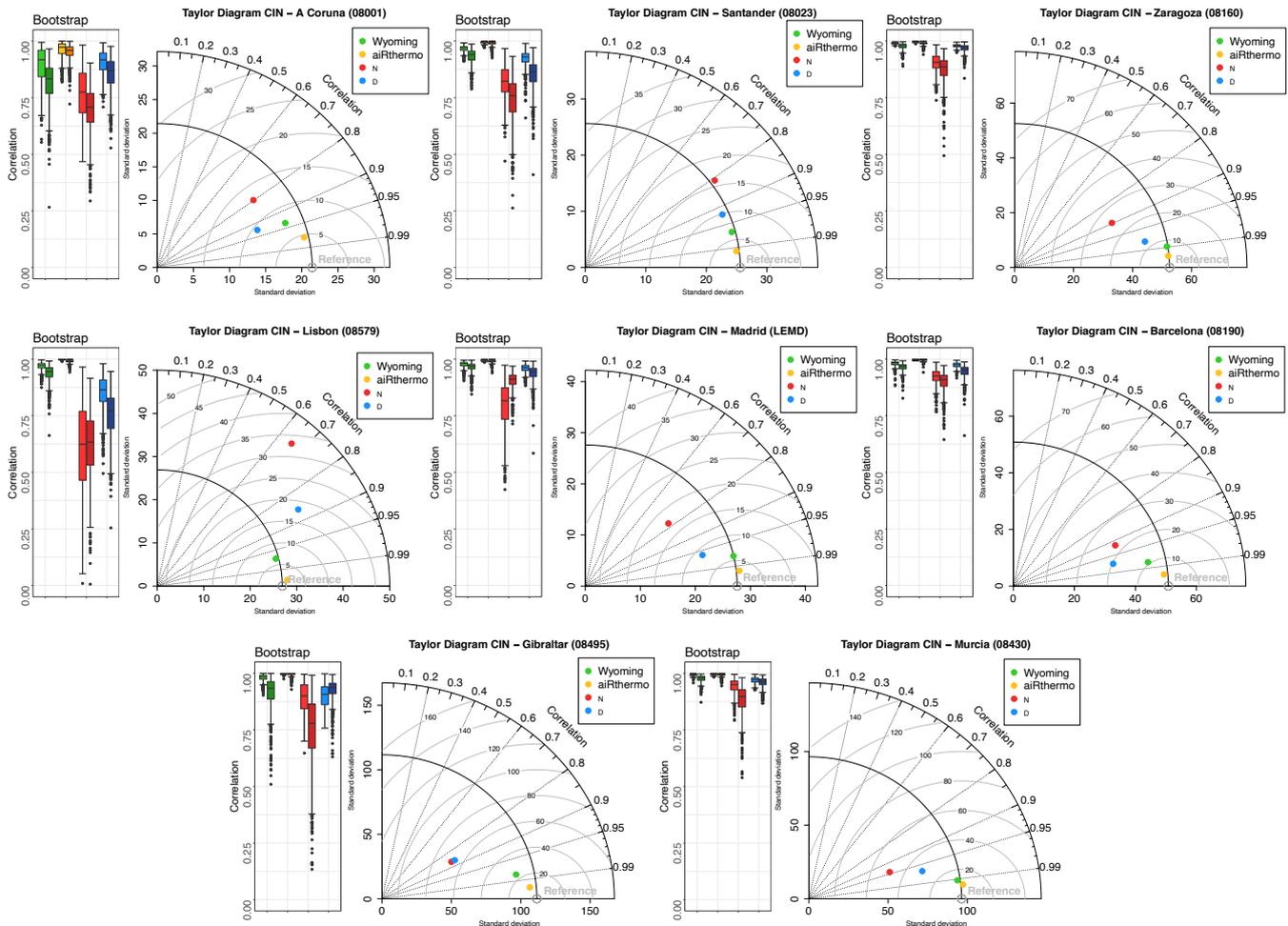


Figure 4. Same as Figures 2 and 3 but for CIN.

The bootstrap analysis presents the same results as for ~~TT and CAPE (Figures 2 and CAPE (Figure 3)~~. However, for Gibraltar, as shown in the Taylor diagram, both WRF experiments produce similar Pearson's correlation values during the bootstrap. If we consider, instead, Spearman's correlations, the worsening of the values is perceptible in A Coruna, Santander, Murcia and Gibraltar. However, in Gibraltar, differences between both WRF experiments arise: WRF D obtained better correlations than WRF N as in the other stations. In contrast to previous results, the poorest correlations for CIN are obtained in stations located in the Atlantic coast as Lisbon and A ~~Coruña~~Coruna.

As for CAPE, the differences between aiRthermo and ~~the reference data Wyoming~~ are highlighted here. This result supports the idea that small differences in the initial conditions of the lifted air parcel (and the determination of the LCL due to differences in the dew point temperature) can trigger large differences in the values of CIN even if the same values of temperature, mixing ratio and pressure are used for its calculation. Again, the differences between both WRF experiments are important and the experiment including data assimilation (D) presents generally closer results to the observed ones.

The seasonal analysis of the ~~four-five~~ datasets (Reference, Wyoming, aiRthermo, N and D) for TT index is presented in Figure 5. In this case, Wyoming, aiRthermo and D are able to correctly simulate the reference seasonal variability of TT index in all the stations and all the seasons. However, N tends to overestimate the ~~variability values of TT~~ in every season and for most of the stations over the IP.

A ~~Coruña~~Coruna and Santander present the largest values during winter, which ~~agrees with the fact that the northern and northwestern IP receive great amounts of rain during~~ is in concordance with the regions where more precipitation is accumulated over the IP in that season (Rodríguez-Puebla et al., 1998; Esteban-Parra et al., 1998; Romero et al., 1999; Iturrioz et al., 2007). Higher values than in winter are observed during spring, but the maximum is recognizable in Madrid. This station is the only one located over central IP and it ~~highlights is in concordance with~~ the maximum in precipitation in that region during that season (Tullot, 2000). However, the other stations also present values above 38 °C. During summer, central, eastern and southern stations (Madrid, Barcelona, Zaragoza and Murcia) are the ones presenting higher values. In that season, the Atlantic stations ~~and Gibraltar (A Coruna, Santander and Gibraltar) and Gibraltar~~ present values below 40 °C. ~~This feature~~ The values of TT in summer on those stations are shorter than the ones in winter, which occurs mainly due to the combined effect of the high increasing values of temperature at 850 and 500 hPa (about 15 and 10 degrees respectively) and the shorter increase of dew point temperature (only a few degrees) in those regions from winter to summer (see Figures A1, A2 and A3 from the Appendix). This pattern is also in agreement with previous studies highlighting that precipitation is important ~~in those regions near the Mediterranean coast~~ during that season (Rodríguez-Puebla et al., 1998; Esteban-Parra et al., 1998; Romero et al., 1999; Iturrioz et al., 2007). Finally, all the stations show similar values in autumn, with the exception of Gibraltar where the values are smaller.

The seasonal analysis for CAPE is presented in Figure 6, and it highlights the spatial and temporal heterogeneity of the areas where ~~convective precipitation can be triggered unstable air masses can be observed~~ over the IP, as also shown by Holley et al. (2014). aiRthermo is Wyoming and aiRthermo are able to reproduce (as expected) the variability of the reference values, and D is able to capture the spread of the values in most of the stations during winter, summer and autumn. However, ~~it overestimates both WRF experiments (particularly D) overestimate~~ CAPE in most of the stations in spring ~~.The other WRF experiment due~~

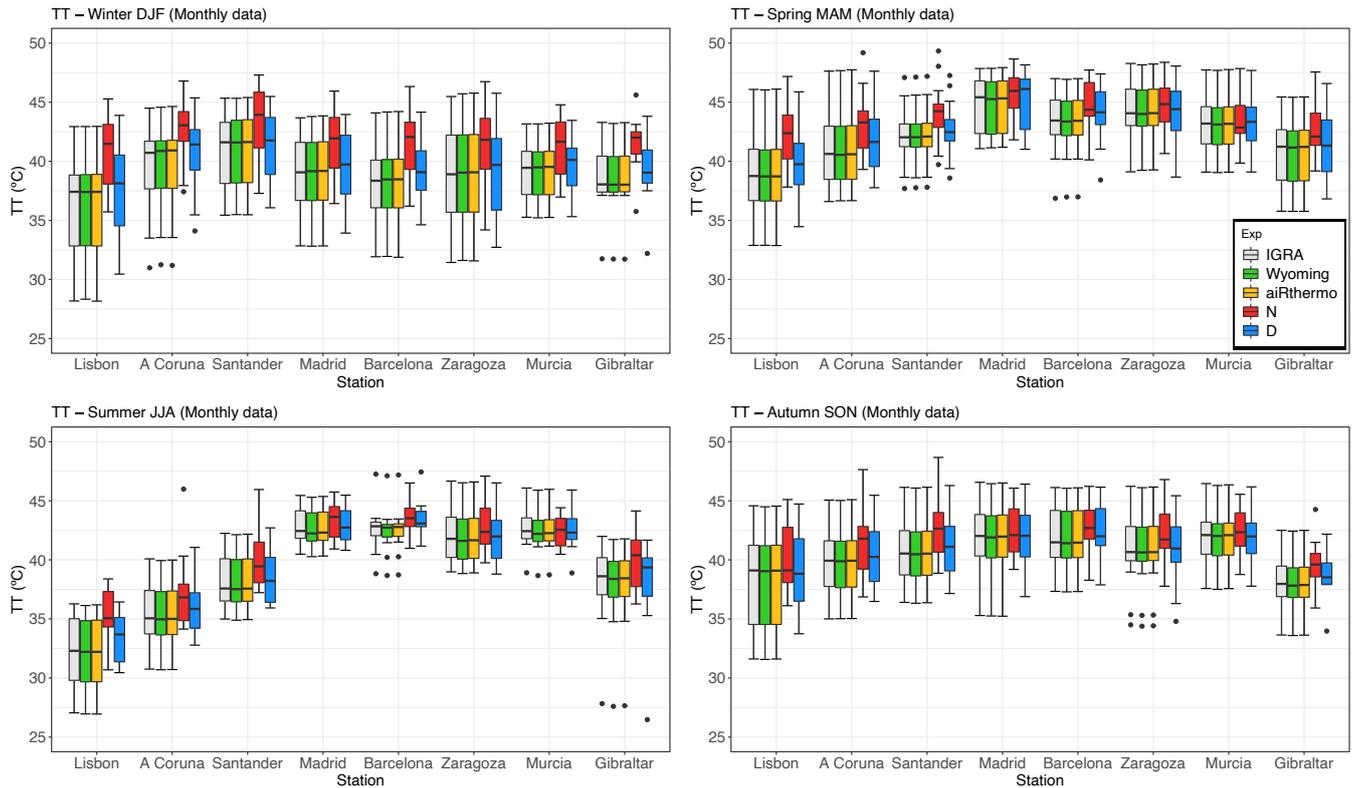


Figure 5. TT index for the reference data (magenta), Wyoming (green), aiRthermo (orange), N (red) and D (blue) computed at each station for every season: winter (top panel, left), spring (top panel, right), summer (bottom panel, left) and autumn (bottom panel, right).

to the differences in the virtual temperature in lower levels compared to reference data (colder near surface and warmer near 800 hPa) and with lifted trajectories for D slightly warmer than the reference ones and N (see Figure A4).

The experiment without data assimilation (N) tends to overestimate CAPE in winter and underestimate it in summer. In winter, this overestimation is caused mainly by colder soundings and warmer lifted trajectories (particularly in Lisbon, A Coruna and Santander - See Figure A5). On the contrary, in summer, the underestimations are caused by warmer sounding levels compared to Reference, which produces that the lifted trajectory crosses earlier than D the sounding, and consequently, underestimating CAPE (particularly in Barcelona, Murcia and Gibraltar - See Figure A6). During spring and autumn, its the underestimations or overestimations of N depend on the station and a clear pattern is not observed.

The lowest values of CAPE are obtained during winter (below 50 J/kg in all the stations), and the largest ones are observed in summer (reaching 500 J/kg in some stations). However, as stated before, the distribution of CAPE is not homogeneous and some regions are more important than others - different regions are prone to higher values during each season. During winter, the three Atlantic stations (A Coruña, Coruna, Santander and Lisbon) and Gibraltar present the highest values of CAPE over the IP. In general, the values are below 50 J/Kg, but some events can get CAPEs over 100 J/kg. During spring, the distribution

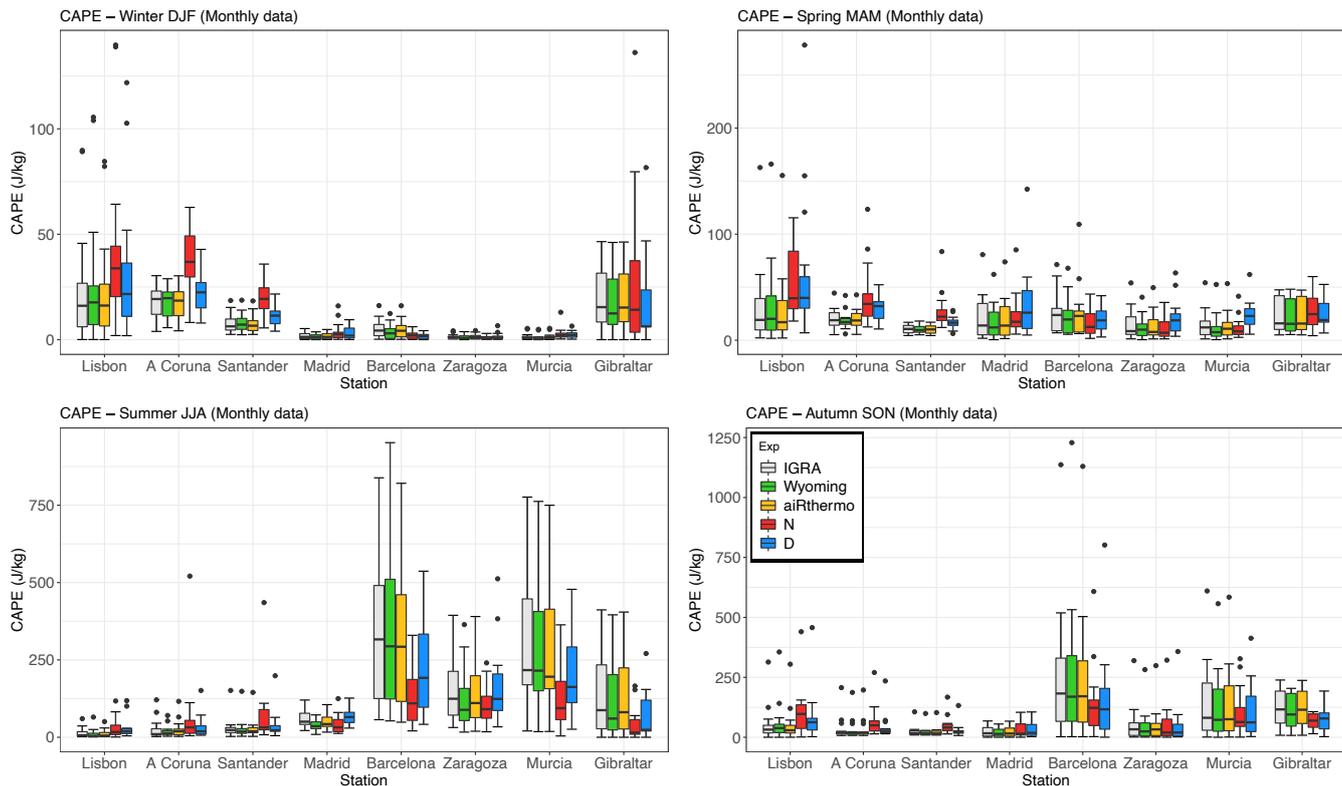


Figure 6. Same as Figure 5 but for CAPE.

of CAPE is quite homogeneous over the IP and only stations such as Lisbon, Madrid or Gibraltar present slightly higher values of CAPE than the other stations. In summer, only the stations located in the eastern and southern parts of the IP present remarkable values of CAPE. Particularly, the most active ones are those located in the Mediterranean coast (Barcelona and Murcia). Finally, during autumn, the regions with high CAPEs are extended towards the inland of the IP, such as Madrid and Zaragoza. During this season, some extreme events can reach values over 1000 J/kg over the Mediterranean coast. This feature was already observed by Siedlecki (2009).

Finally, the seasonal analysis for CIN is presented in Figure 7, and it highlights the stations where the inhibition is important. In general, ~~aiRthermo tends to overestimate the observed variability~~ Wyoming tends to underestimate the values of CIN in most of the stations and in every season. ~~On the contrary, both~~, while aiRthermo is able to capture it. Both WRF simulations (but particularly the experiment without data assimilation) tend to underestimate the observed variability.

The values of CIN are smaller in winter and spring, and the maximum is observed in summer. During winter, CIN is higher in Gibraltar and in the Atlantic stations (Lisbon, A ~~Coruña~~ Coruña and Santander) than in the other stations from the IP. However, these values are small compared to those from other seasons. During spring, the values are higher than in winter, and similar values are observed in most of the stations (around 10 J/kg), with the exception of Barcelona where the CIN reaches

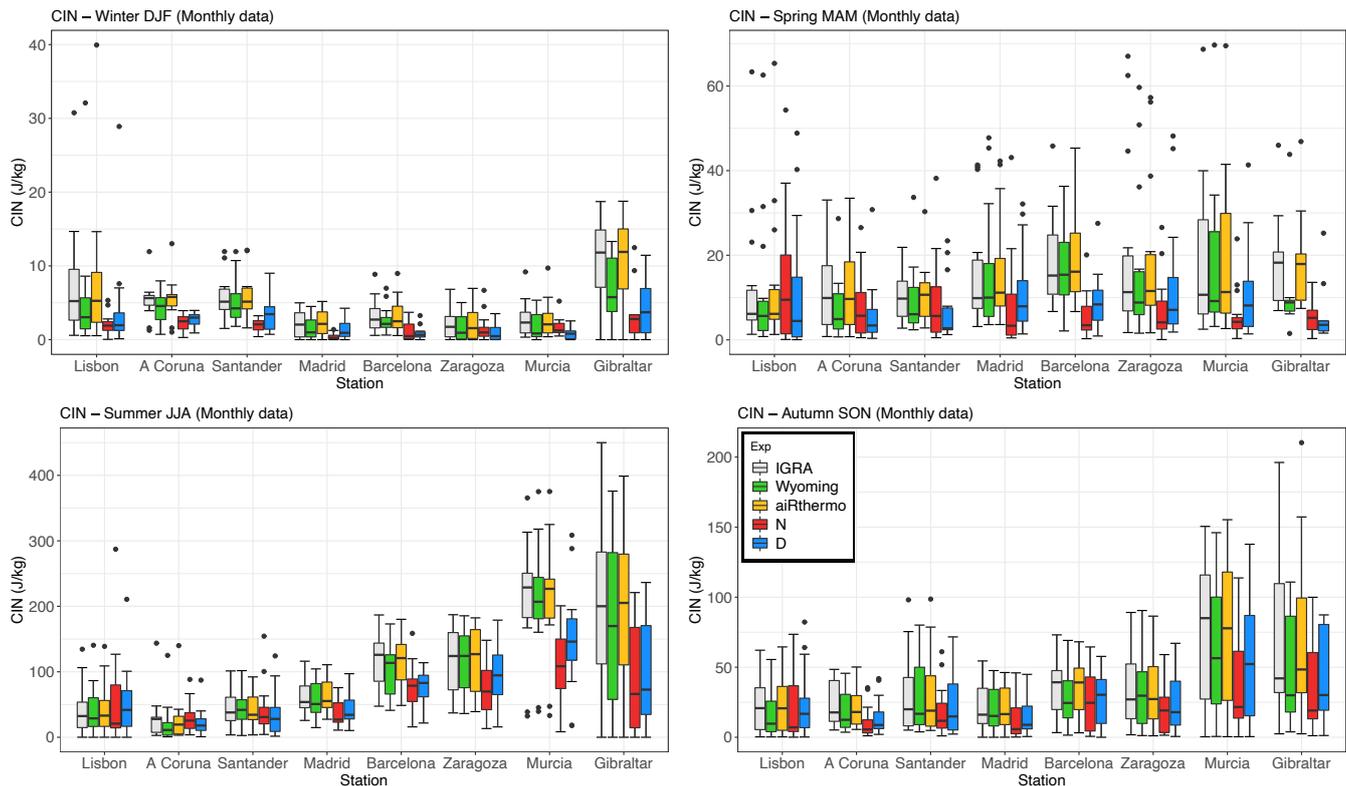


Figure 7. Same as Figures 5 and 6 but for CIN.

values of 20 J/kg. In summer, the values are higher in all the stations, but particularly in those from the eastern and southern IP (Barcelona, Zaragoza, Murcia and Gibraltar). The same regime is observed for autumn, but the values are smaller than in summer. These values during summer and autumn are in agreement with Siedlecki (2009), who found CIN means above 100 J/Kg in the west Mediterranean sea and surrounding countries.

390 As stated before, in the final phase of this study, the same procedure for the calculation of the instability indices at each station was extended to each grid point included in the IP. The mean winter and summer spatial patterns at 00 and 12 UTC were calculated for both WRF experiments. [The spatial distribution for TT is shown in Figure 8](#) ~~shows the results for TT. The~~ [, which highlights the](#) heterogeneity of the results ~~is highlighted in those results~~. The differences between both simulations are ~~remarkable~~ [observable](#), but also those between day and night. Additionally, it can be seen that TT cannot be calculated in most
395 of the mountain regions of the IP because the 850 hPa layer is near the surface or below ground.

During winter, the TT maps show that N yields higher values than D, which is in agreement with the overestimation observed in Figure 5. At 00 UTC, according to D, the regions where ~~thunderstorms are likely to happen~~ [unstable air masses are observed](#) are those in the Cantabrian coast and in the southeastern IP. Both regions are surrounded by ~~important~~ [remarkable](#) mountainous systems such as the Cantabrian Range and the Baetic system, [so the lifting that can trigger convection can appear](#). For N,

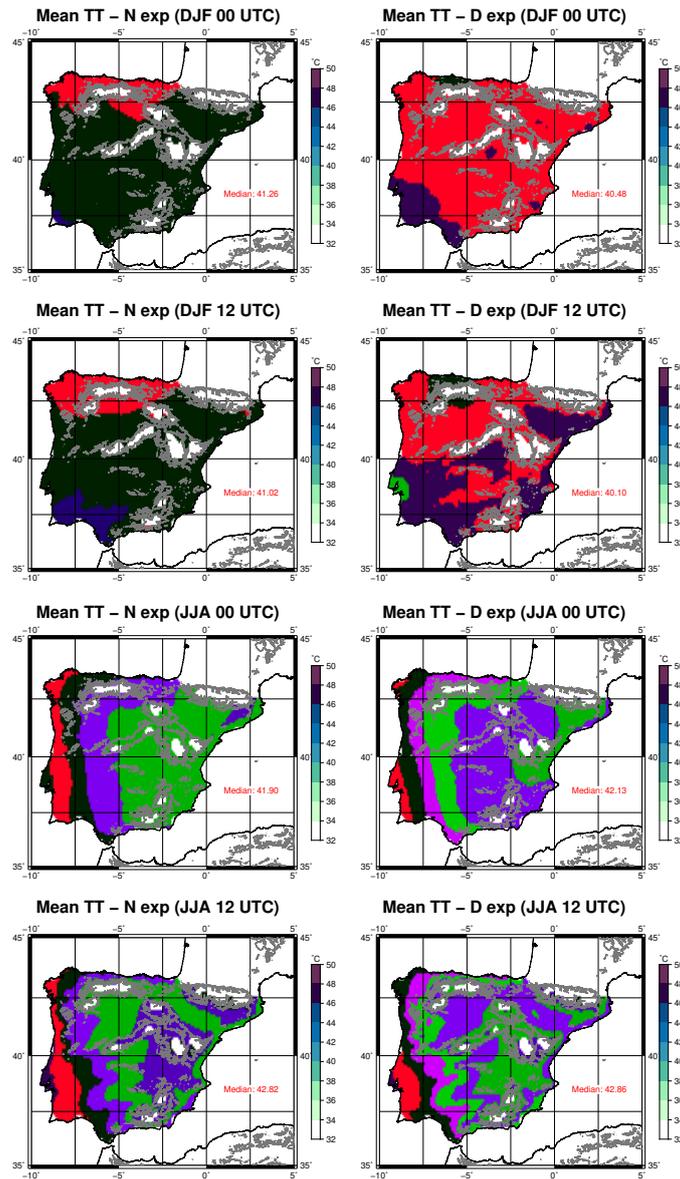


Figure 8. Spatial distribution of mean TT for period 2010-2014 over the IP as computed from N (first column) and D (second column) for winter (rows 1 and 2) and summer (rows 3 and 4) at 00 (rows 1 and 3) and 12 UTC (rows 2 and 4). The median value (°C) of each map is presented in the bottom right corner of the plots.

400 the possibility of thunderstorms is these areas are also extended to the rest of the IP, with the exception of the southwestern corner where the values are small. At 12 UTC, after solar irradiance has started heating up the land, the areas with high chance for thunderstorms regions extend towards inland areas. In the experiment with data assimilation (D), most of the northern plateau presents high values of TT, and the lowest values are observed near the coastal valleys of the southwestern corner

and the Mediterranean coast (like the Ebro basin or Murcia). In the case of N, the lowest values are observed mainly in the southwestern IP near the Guadalquivir valley. According to Figures A1, A2 and A3, the lowest values of TT observed near the coastal valleys of both WRF experiments are originated due to the low values for dew point temperature observed in those regions, which at the same time are originated by low mixing ratio values.

As in Figure 5, much higher TT values are obtained during summer over the IP, particularly at 12 UTC. At 00 UTC, a west-east gradient is observed in both WRF simulations. However, the values depicted in the Mediterranean coast are higher for the experiment including data assimilation (D). At 12 UTC, the ~~areas with higher chance to develop thunderstorms regions~~ with unstable air masses extend towards the central area. In this case, ~~the regions they~~ are located near the mountains of the IP. Particularly, in the southern slope of the Pyrenees and in the ~~proximities proximity~~ of the Iberian, ~~central~~ Central and Baetic systems. The minimum TTs are observed in the western part of the IP, but particularly near Lisbon. The intensity of the most extreme values of TT is higher in D (with data assimilation).

The maps for CAPE at 00 and 12 UTC during winter and summer are presented in Figure 9. During winter, as shown in Figure 6, the N experiment presents higher values than D. At 00 UTC, the patterns are really similar for both WRF experiments. The main difference between them is observed in the western Atlantic coast of the IP, where higher CAPE values are obtained for D. At 12 UTC, the unstable ~~area is air masses are~~ found in the western coast of the IP in both simulations, and ~~it is extended~~ they extend further inland than at 00 UTC, particularly near the Tagus and Guadalquivir rivers. Again, the values are higher for N, but the pattern is similar in both experiments.

~~On the contrary~~ Compared to what is observed during winter, CAPE is higher during summer for the experiment including data assimilation. This is in agreement with the station analysis from previous Figure 6. At 00 UTC, the area ~~where thunderstorms can be developed~~ with higher CAPE is observed in the northern and eastern IP, but particularly near the Mediterranean coast. However, at 12 UTC, this area with high values (over 250 J/kg) extends towards the interior and in the experiment including data assimilation it also covers the southern part of the Pyrenees. Additionally, high values are observed in most of the IP (except the southwestern corner for N), but particularly in the simulation including data assimilation.

Finally, regarding CIN, the maps for the mean values at 00 and 12 UTC during winter and summer are presented in Figure 10. In reverse to what we found for CAPE, CIN is usually higher at 00 UTC than at 12 UTC (with the exception of Murcia in summer at 12 UTC). During winter, at 00 UTC, both simulations show small values over the IP, and only some high values are observed in the western and southwestern corner of the IP (and particularly for N). At 12 UTC, the areas are confined to those coastal regions, but they also extend to the Mediterranean coast in the D experiment.

During summer, at 00 UTC, the most remarkable values are obtained in both simulations along the Ebro basin and near the Mediterranean coast. However, the CIN inland is higher for D. At 12 UTC, less inhibition is observed in the eastern valleys of the IP (with the exception of Murcia, which presents extremely high values of CIN). ~~In~~ At the same time, an increase in the convective inhibition over the Guadalquivir basin is shown. The extension towards the interior is again higher for D (including data assimilation).

~~Thus, some clear dynamics~~ Comparing the results shown in Figures 5 and 6 (or in Figures 8 and 9), it seems that there is a discrepancy between the results from TT and CAPE since maximal values of these indices are not observed in the same

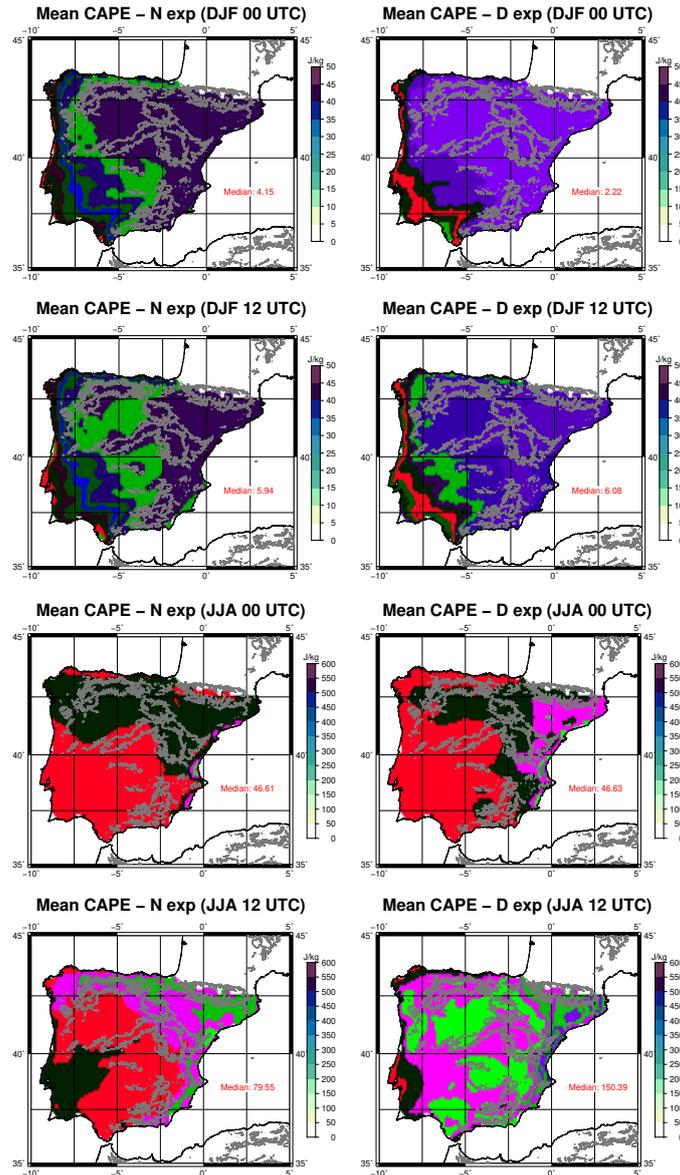


Figure 9. Same as Figure 8 but for CAPE.

440 regions. However, it must be taken into account that these results for CAPE (and CIN) are obtained from the entire series of 12 hourly values obtained during 2010-2014. Thus, these values are not restricted to highly convective events, and consequently, Figures 6 and 7 (also Figures 9 and 10) must be compared in combination to the values of TT. Additionally, TT and CAPE are indices for atmospheric instability, but they are not related through a simple linear relationship (R^2 below 0.2 for all the stations and seasons), particularly for stable or neutral atmospheres. Thus, since CAPE and CIN are dependent on the entire profile of the atmosphere, they should be considered more reliable than TT, which only takes into account two pressure levels.

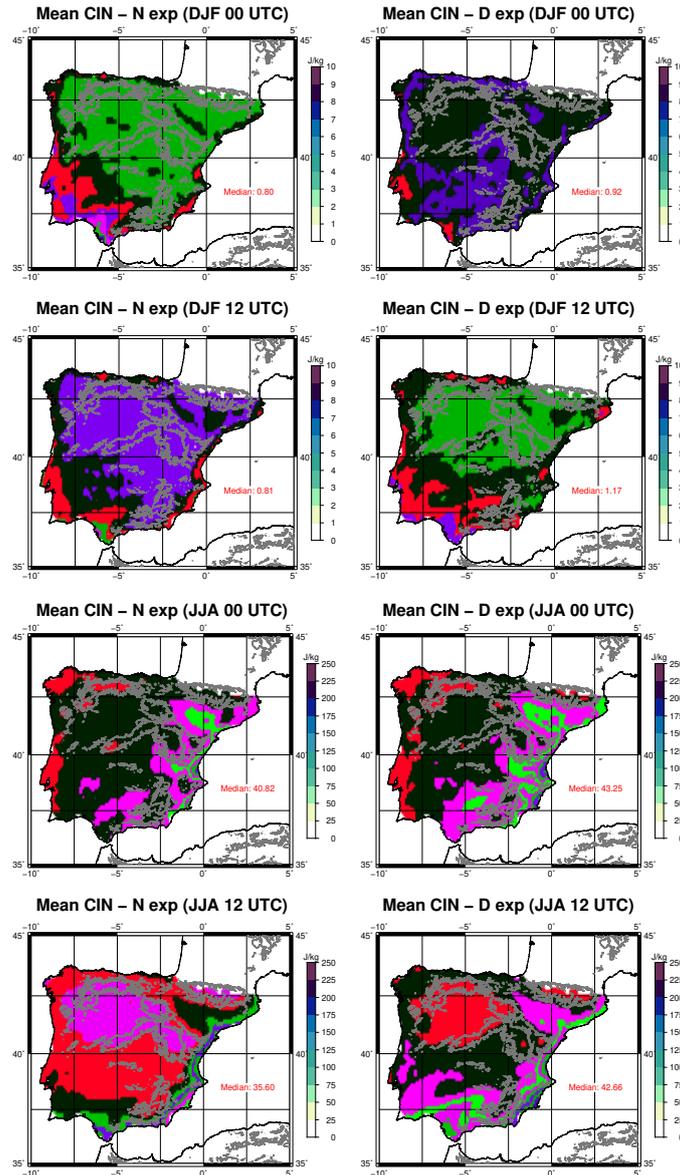


Figure 10. Same as Figures 8 and 10 but for CIN.

445 Taking in consideration the information presented above, some clear patterns arise from these results. During winter, the areas where thunderstorms can be developed with unstable air masses are located in the Atlantic coast of the IP, and they are more active the instability is more intense during the afternoon as CIN is really high in those region until 12 UTC. However, during summer, the unstable areas are located to the north of the Mediterranean coast. Thunderstorms will tend to occur The intensity increases during the afternoon in those regions, but they can also develop some unstable areas can also appear
 450 before 12 UTC even if the inhibition is high during that period (CAPE is also high). These features are in agreement with

the precipitation dynamics observed in previous studies over the IP (Rodríguez-Puebla et al., 1998; Esteban-Parra et al., 1998; Romero et al., 1999; Iturrioz et al., 2007). The patterns for CAPE observed during winter and summer are similar to those obtained by the regional analysis performed by Viceto et al. (2017). However, their values are comparable to those obtained by our experiment without data assimilation (N). In this case, the data assimilation (D) produces higher values, but much more realistic than the ones from those simulations (N) without it (according to Figure 6).

4 Conclusions

The main purpose of this paper is to evaluate the ability to simulate ~~the instability conditions that can trigger thunderstorms~~ unstable conditions over the IP with two high-resolution simulations created with the state-of-the-art model WRF. One of these simulations is driven by the boundary conditions provided by ERA-Interim reanalysis (N experiment), while the second one presents the same configuration but including the additional 3DVAR data assimilation step every 6 hours (D experiment). Three instability indices were evaluated: TT index, CAPE and CIN. All of them were calculated from the outputs of the model using the publicly available R package *aiRthermo*, also developed by the authors.

In order to validate these indices, their values were downloaded from the University of Wyoming server for the eight radiosondes available over the IP. Additionally, ~~pressure, temperature, potential temperature, temperature,~~ mixing ratio and height at all the available pressure levels from the radiosondes were also retrieved ~~from the server.~~ In that case, also from IGRA from NOAA. These variables were used to calculate with *aiRthermo* again these three indices. Comparing these new values with the ones retrieved directly from the University of Wyoming, the small differences which can only be attributed to different methodologies can be obtained.

First, the correlation, SD and RMSE were computed for each of the indices for both simulations and for the ones calculated with the data at pressure levels from the University of Wyoming following our own methodology. The ~~three~~ four of them were compared against the reference values (the ones ~~downloaded from the University of Wyoming~~ from IGRA) by means of Taylor diagrams. According to these results, small differences can be observed due to the different ~~methodologies~~ methods used for the calculation of the indices (particularly associated to the use of measured or calculated dew point temperature). However, these differences are more important for CAPE and CIN than for TT because they are highly dependant on the initial conditions for the calculation of the vertical integrations, while TT index is ~~not only dependant on two discrete pressure levels.~~ Between both WRF simulations, the most accurate results are produced by the experiment with data assimilation (D). The bootstrap analysis with resampling also supports this result.

Then, the seasonal analysis was carried out for each index. For TT, the differences between methods are really small in every season. Between both WRF experiments, N tends to overestimate the reference variability, while D is able to capture it. In the case of CAPE and CIN, the differences between methods are larger, but not as those within both WRF experiments. However, D is able to produce closer values to the reference ~~values~~ than N. During winter, the unstable ~~areas where thunderstorms can be triggered~~ air masses are located mainly in over the stations from the Atlantic coast of the IP. All the stations are quite active during spring. However, the regime changes to the Mediterranean coast during summer, and also autumn but with less intensity.

Finally, the calculation of the indices was also carried out for every grid point over the IP in both WRF simulations, particularly for winter and summer and at 00 and 12 UTC. All the three indices agree highlighting the heterogeneity of the patterns observed. The D experiment, which is the most accurate one according to the previous analysis, shows that during winter the convective activity is unstable areas are found along the entire Atlantic coast, but particularly in the southwestern corner of the IP when the instability is extended towards inland regions. During summer, this feature is reversed, and the region most prone to intense convection develop unstable air masses is located in the Ebro basin and the Mediterranean coast. The possibility of a storm developing inhibition (high values of CIN) is strong at 00 UTC is rather small because of the high values of CIN in those regions, but that is highly reduced at 12 UTC.

Data availability. These results can be reproduced using the postprocessed outputs from the model available in <https://doi.org/10.5281/zenodo.3611343>.

Appendix A: [Supplementary figures](#)

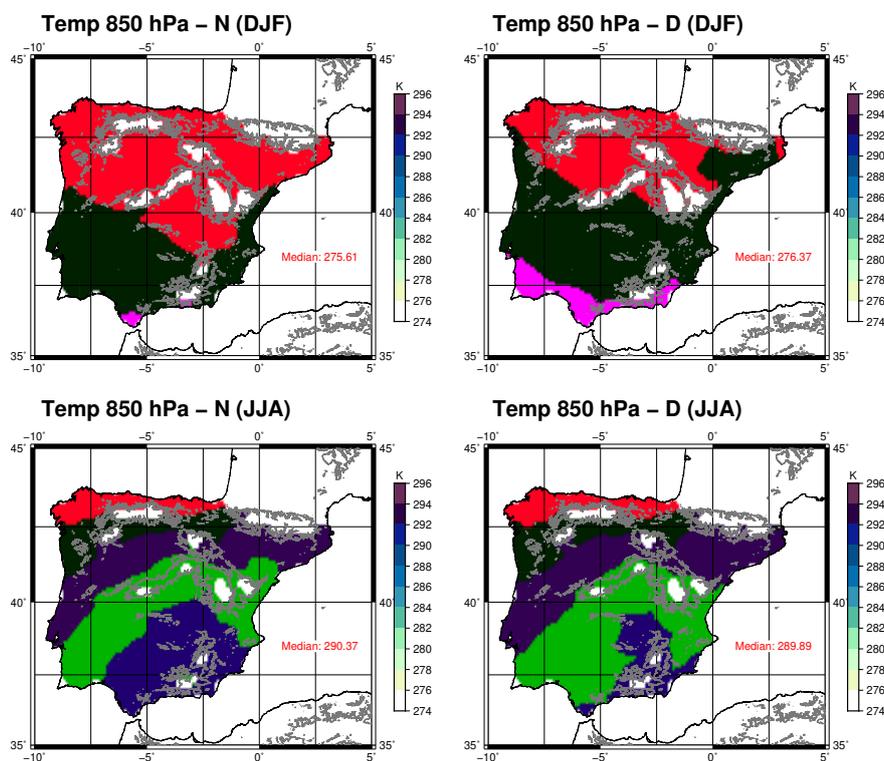


Figure A1. [Spatial distribution of mean temperature at 850 hPa for period 2010-2014 over the IP as computed from N \(first column\) and D \(second column\) for winter and summer. The median value \(K\) is in the bottom right corner of the plots.](#)

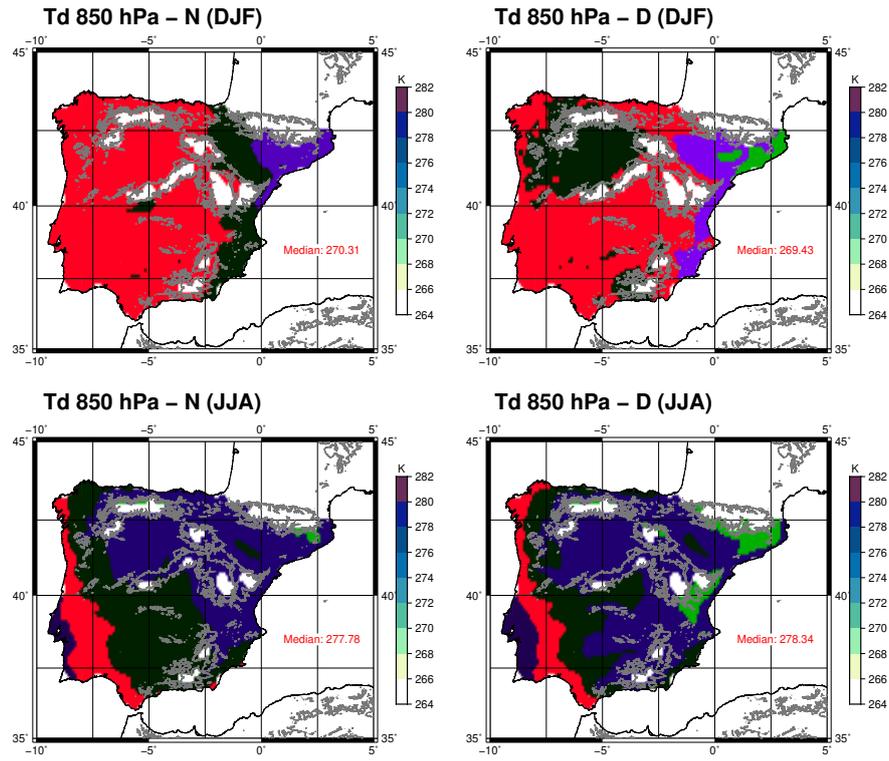


Figure A2. Same as Figure A1 but for dew point temperature at 850 hPa.

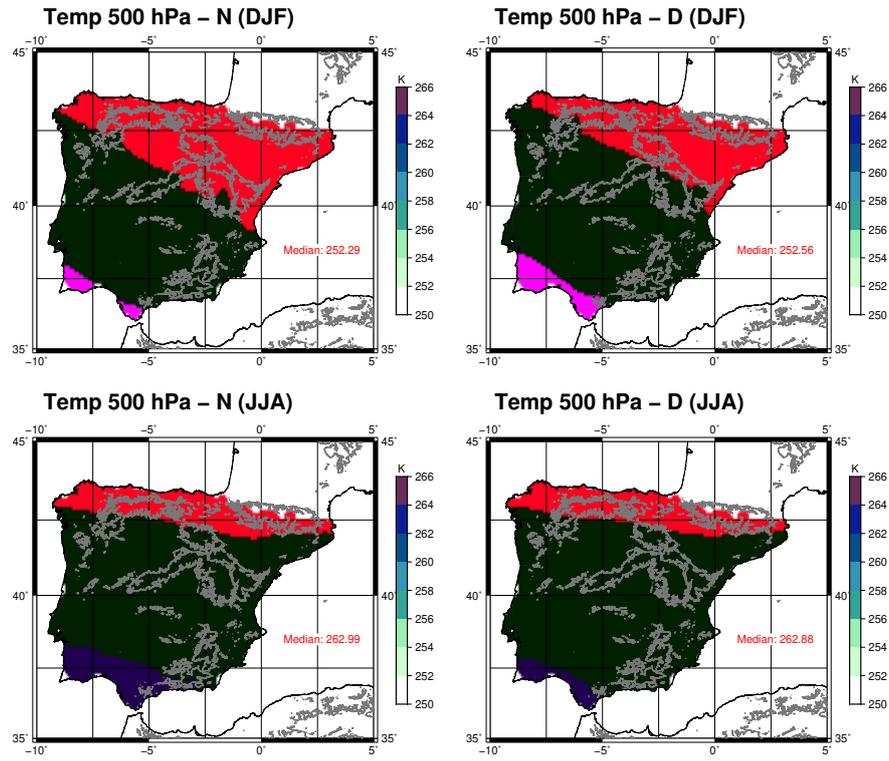


Figure A3. Same as Figures A1 and A2 but for temperature at 500 hPa.

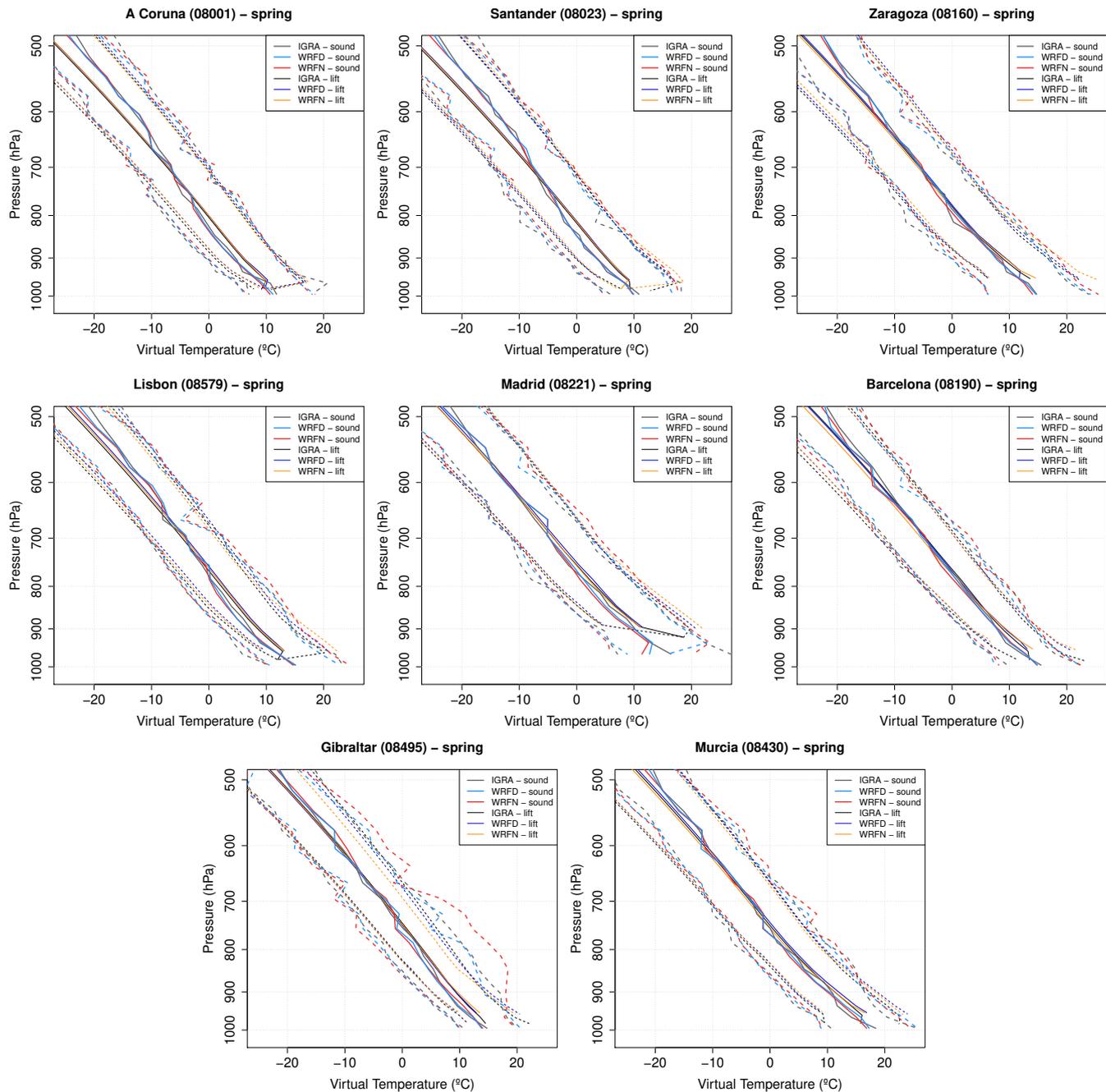


Figure A4. Vertical profiles of virtual temperature for the sounding levels and for the lifted parcel during spring. The dashed lines represent the 5 and 95 percentiles, and the solid lines the median.

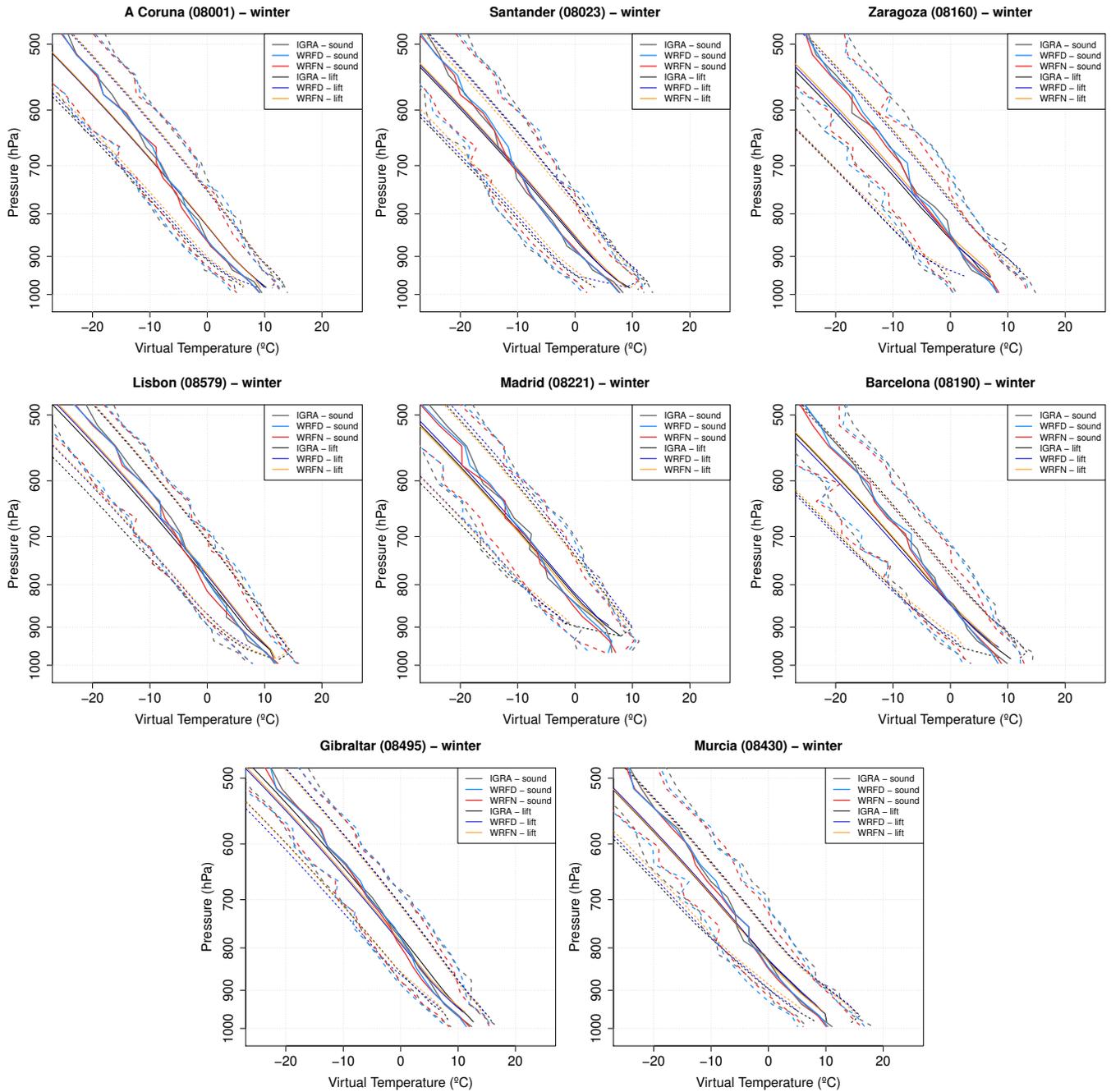


Figure A5. Same as Figure A4 but for winter.

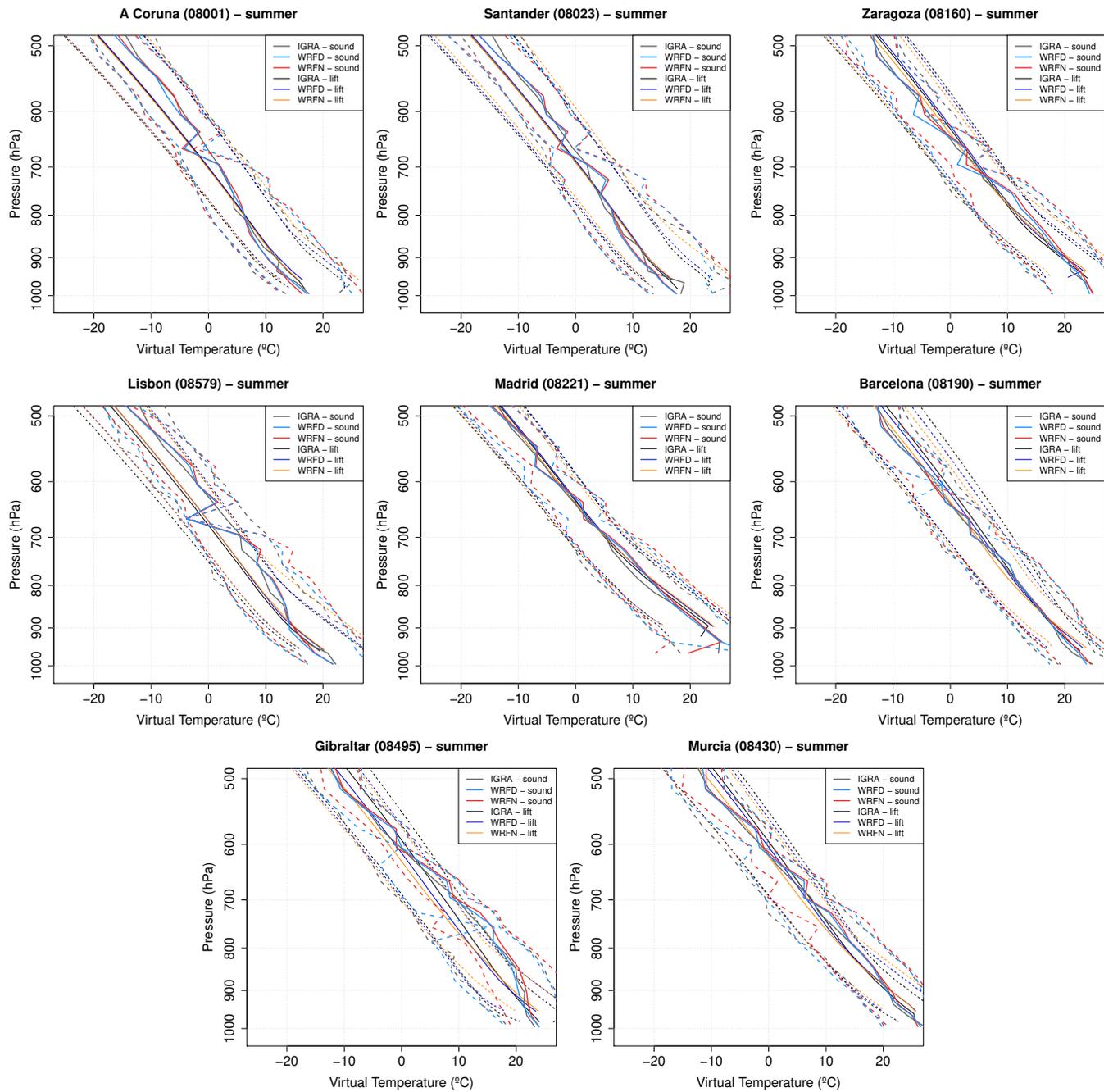


Figure A6. Same as Figures A4 and A5 but for summer.

495 *Author contributions.* The methodology and the software was developed by S.J.G.-R., S.C.-M. and J.S.; The conceptualization, preparation of datasets and analysis was carried out by all the authors; The original draft of the paper was written by S.J.G.-R., but all the authors took part in the edition and revision of it.

Competing interests. The authors declare no conflict of interest.

500 *Acknowledgements.* S.J.G.R. is now funded by the Oeschger Centre for Climate Change Research (OCCR), but during his PhD he was supported by a FPI Predoctoral Research grant (MINECO, BES-2014-069977). This study was also supported through MINECO project CGL2016-76561-R from the Spanish Government (MINECO/ERDF, UE) and the grant GIU 17/002 from the University of the Basque Country. The computational resources were provided by I2BASQUE, and the authors thank the creators of WRF/ARW and WRFDA systems. [Authors thank the anonymous reviewers for their comments, which have helped to improve the paper.](#) Finally, most of the calculations were carried out with R ([R Core Team, 2018](#)), and the authors want to thank all the authors of the packages used for it.

505 References

- Alexander, G. D. and Young, G. S.: The Relationship between EMEX Mesoscale Precipitation Feature Properties and Their Environmental Characteristics, *Monthly Weather Review*, 120, 554–564, [https://doi.org/10.1175/1520-0493\(1992\)120<0554:TRBEMP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1992)120<0554:TRBEMP>2.0.CO;2), 1992.
- Angus, P., Rasmussen, S., and Seiter, K.: Short-term prediction of thunderstorm probability and intensity by screening observational and derived predictors, in: Reprints AMS 15th Conference on Severe Local Storms, Baltimore, MA, pp. 368–371, 1988.
- 510 Argüeso, D., Hidalgo-Muñoz, J. M., Gámiz-Fortis, S. R., Esteban-Parra, M. J., Dudhia, J., and Castro-Díez, Y.: Evaluation of WRF Parameterizations for Climate Studies over Southern Spain Using a Multistep Regionalization, *Journal of Climate*, 24, 5633–5651, <https://doi.org/10.1175/JCLI-D-11-00073.1>, 2011.
- Barker, D., Huang, X.-Y., Liu, Z., Auligné, T., Zhang, X., Rugg, S., Ajjaji, R., Bourgeois, A., Bray, J., Chen, Y., Demirtas, M., Guo, Y.-R., Henderson, T., Huang, W., Lin, H.-C., Michalakes, J., Rizvi, S., and Zhang, X.: The Weather Research and Forecasting Model's
515 Community Variational/Ensemble Data Assimilation System: WRFDA, *Bulletin of the American Meteorological Society*, 93, 831–843, <https://doi.org/10.1175/BAMS-D-11-00167.1>, 2012.
- Barker, D. M., Huang, W., Guo, Y.-R., Bourgeois, A. J., and Xiao, Q. N.: A Three-Dimensional Variational Data Assimilation System for MM5: Implementation and Initial Results, *Monthly Weather Review*, 132, 897–914, [https://doi.org/10.1175/1520-0493\(2004\)132<0897:ATVDAS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<0897:ATVDAS>2.0.CO;2), 2004.
- 520 Blanchard, D. O.: Assessing the Vertical Distribution of Convective Available Potential Energy, *Weather and Forecasting*, 13, 870–877, [https://doi.org/10.1175/1520-0434\(1998\)013<0870:ATVDOC>2.0.CO;2](https://doi.org/10.1175/1520-0434(1998)013<0870:ATVDOC>2.0.CO;2), 1998.
- Brooks, H.: Severe thunderstorms and climate change, *Atmospheric Research*, 123, 129 – 138, <https://doi.org/https://doi.org/10.1016/j.atmosres.2012.04.002>, 2013.
- Brooks, H. E., Lee, J. W., and Craven, J. P.: The spatial distribution of severe thunderstorm and tornado environments from global reanalysis
525 data, *Atmospheric Research*, 67-68, 73 – 94, [https://doi.org/https://doi.org/10.1016/S0169-8095\(03\)00045-0](https://doi.org/https://doi.org/10.1016/S0169-8095(03)00045-0), 2003.
- Correoso, J. F., Hernández, E., García-Herrera, R., Barriopedro, D., and Paredes, D.: A 3-year study of cloud-to-ground lightning flash characteristics of Mesoscale convective systems over the Western Mediterranean Sea, *Atmospheric Research*, 79, 89 – 107, <https://doi.org/https://doi.org/10.1016/j.atmosres.2005.05.002>, 2006.
- Craven, J. P., Jewell, R. E., and Brooks, H. E.: Comparison between Observed Convective Cloud-Base Heights and Lift-
530 ing Condensation Level for Two Different Lifted Parcels, *Weather and Forecasting*, 17, 885–890, [https://doi.org/10.1175/1520-0434\(2002\)017<0885:CBOCCB>2.0.CO;2](https://doi.org/10.1175/1520-0434(2002)017<0885:CBOCCB>2.0.CO;2), 2002.
- Dai, A.: Recent changes in the diurnal cycle of precipitation over the United States, *Geophysical Research Letters*, 26, 341–344, <https://doi.org/10.1029/1998GL900318>, 1999.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer,
535 P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597, <https://doi.org/10.1002/qj.828>, 2011.
- 540 DeRubertis, D.: Recent Trends in Four Common Stability Indices Derived from U.S. Radiosonde Observations, *Journal of Climate*, 19, 309–323, <https://doi.org/10.1175/JCLI3626.1>, 2006.

- Diffenbaugh, N. S., Scherer, M., and Trapp, R. J.: Robust increases in severe thunderstorm environments in response to greenhouse forcing, *Proceedings of the National Academy of Sciences*, 110, 16 361–16 366, <https://doi.org/10.1073/pnas.1307758110>, 2013.
- 545 Doswell, C. A. and Rasmussen, E. N.: The Effect of Neglecting the Virtual Temperature Correction on CAPE Calculations, *Weather and Forecasting*, 9, 625–629, [https://doi.org/10.1175/1520-0434\(1994\)009<0625:TEONTV>2.0.CO;2](https://doi.org/10.1175/1520-0434(1994)009<0625:TEONTV>2.0.CO;2), 1994.
- Doswell, C. A., Ramis, C., Romero, R., and Alonso, S.: A Diagnostic Study of Three Heavy Precipitation Episodes in the Western Mediterranean Region, *Weather and Forecasting*, 13, 102–124, [https://doi.org/doi:10.1175/1520-0434\(1998\)013<0102:ADSOTH>2.0.CO;2](https://doi.org/doi:10.1175/1520-0434(1998)013<0102:ADSOTH>2.0.CO;2), 1998.
- Efron, B. and Gong, G.: A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation, *The American Statistician*, 37, 36–48, <https://doi.org/10.1080/00031305.1983.10483087>, 1983.
- 550 Enno, S.-E., Sugier, J., Alber, R., and Seltzer, M.: Lightning flash density in Europe based on 10 years of ATDnet data, *Atmospheric Research*, 235, 104 769, <https://doi.org/https://doi.org/10.1016/j.atmosres.2019.104769>, 2020.
- Eshel, G. and Farrell, B. F.: Thermodynamics of Eastern Mediterranean Rainfall Variability, *Journal of the Atmospheric Sciences*, 58, 87–92, [https://doi.org/10.1175/1520-0469\(2001\)058<0087:TOEMRV>2.0.CO;2](https://doi.org/10.1175/1520-0469(2001)058<0087:TOEMRV>2.0.CO;2), 2001.
- 555 Esteban-Parra, M. J., Rodrigo, F. S., and Castro-Diez, Y.: Spatial and temporal patterns of precipitation in Spain for the period 1880–1992, *International Journal of Climatology*, 18, 1557–1574, [https://doi.org/10.1002/\(SICI\)1097-0088\(19981130\)18:14<1557::AID-JOC328>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1097-0088(19981130)18:14<1557::AID-JOC328>3.0.CO;2-J), 1998.
- Fernández, J., Montávez, J. P., Sáenz, J., González-Rouco, J. F., and Zorita, E.: Sensitivity of the MM5 mesoscale model to physical parameterizations for regional climate studies: Annual cycle, *Journal of Geophysical Research: Atmospheres*, 112, <https://doi.org/10.1029/2005JD006649>, 2007.
- 560 Galway, J. G.: The lifted index as a predictor of latent instability, *Bulletin of the American Meteorological Society*, 37, 528–529, 1956.
- Gascón, E., Merino, A., Sánchez, J., Fernández-González, S., García-Ortega, E., López, L., and Hermida, L.: Spatial distribution of thermodynamic conditions of severe storms in southwestern Europe, *Atmospheric Research*, 164-165, 194 – 209, <https://doi.org/https://doi.org/10.1016/j.atmosres.2015.05.012>, 2015.
- 565 Gayà, M., Homar, V., Romero, R., and Ramis, C.: Tornadoes and waterspouts in the Balearic Islands: phenomena and environment characterization, *Atmospheric Research*, 56, 253 – 267, [https://doi.org/https://doi.org/10.1016/S0169-8095\(00\)00076-4](https://doi.org/https://doi.org/10.1016/S0169-8095(00)00076-4), 2001.
- George, J. J.: *Weather forecasting for aeronautics*, Academic Press, San Diego, page 411, 1960.
- González-Rojí, S. J., Sáenz, J., Ibarra-Berastegi, G., and Díaz de Argandoña, J.: Moisture balance over the Iberian Peninsula according to a regional climate model: The impact of 3DVAR data assimilation., *Journal of Geophysical Research: Atmospheres*, 123, 708–729, <https://doi.org/10.1002/2017JD027511>, 2018.
- 570 González-Rojí, S. J., Wilby, R. L., Sáenz, J., and Ibarra-Berastegi, G.: Harmonized evaluation of daily precipitation downscaled using SDSM and WRF+WRFDA models over the Iberian Peninsula, *Climate Dynamics*, 53, 1413–1433, <https://doi.org/10.1007/s00382-019-04673-9>, 2019.
- González-Rojí, S. J., Sáenz, J., Díaz de Argandoña, J., and Ibarra-Berastegi, G.: Moisture Recycling over the Iberian Peninsula: The Impact of 3DVAR Data Assimilation, *Atmosphere*, 11, 19, <https://doi.org/10.3390/atmos11010019>, 2020.
- Haklander, A. J. and Van Delden, A.: Thunderstorm predictors and their forecast skill for the Netherlands, *Atmospheric Research*, 67-68, 273 – 299, [https://doi.org/https://doi.org/10.1016/S0169-8095\(03\)00056-5](https://doi.org/https://doi.org/10.1016/S0169-8095(03)00056-5), 2003.
- Holley, D. M., Dorling, S. R., Steele, C. J., and Earl, N.: A climatology of convective available potential energy in Great Britain, *International Journal of Climatology*, 34, 3811–3824, <https://doi.org/10.1002/joc.3976>, 2014.

- 580 Hong, S.-Y., Dudhia, J., and Chen, S.-H.: A Revised Approach to Ice Microphysical Processes for the Bulk Parameterization of Clouds and Precipitation, *Monthly Weather Review*, 132, 103–120, [https://doi.org/10.1175/1520-0493\(2004\)132<0103:ARATIM>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<0103:ARATIM>2.0.CO;2), 2004.
- Huntrieser, H., Schiesser, H. H., Schmid, W., and Waldvogel, A.: Comparison of Traditional and Newly Developed Thunderstorm Indices for Switzerland, *Weather and Forecasting*, 12, 108–125, [https://doi.org/10.1175/1520-0434\(1997\)012<0108:COTAND>2.0.CO;2](https://doi.org/10.1175/1520-0434(1997)012<0108:COTAND>2.0.CO;2), 1997.
- Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-
585 lived greenhouse gases: Calculations with the AER radiative transfer models, *Journal of Geophysical Research: Atmospheres*, 113, <https://doi.org/10.1029/2008JD009944>, 2008.
- Iturrioz, I., Hernández, E., Ribera, P., and Queralt, S.: Instability and its relation to precipitation over the Eastern Iberian Peninsula, *Advances in Geosciences*, 10, 45–50, <https://doi.org/10.5194/adgeo-10-45-2007>, 2007.
- Jerez, S., Montavez, J. P., Gomez-Navarro, J. J., Jimenez, P. A., Jimenez-Guerrero, P., Lorente, R., and Gonzalez-Rouco, J. F.: The role of
590 the land-surface model for climate change projections over the Iberian Peninsula, *Journal of Geophysical Research: Atmospheres*, 117, <https://doi.org/10.1029/2011JD016576>, 2012.
- Johns, R. H. and Doswell, C. A.: Severe Local Storms Forecasting, *Weather and Forecasting*, 7, 588–612, [https://doi.org/10.1175/1520-0434\(1992\)007<0588:SLSF>2.0.CO;2](https://doi.org/10.1175/1520-0434(1992)007<0588:SLSF>2.0.CO;2), 1992.
- Jones, R. G., Murphy, J. M., and Noguera, M.: Simulation of climate change over Europe using a nested regional-climate model. I: Assessment
595 of control climate, including sensitivity to location of lateral boundaries, *Quarterly Journal of the Royal Meteorological Society*, 121, 1413–1449, <https://doi.org/10.1002/qj.49712152610>, 1995.
- Kaltenböck, R., Diendorfer, G., and Dotzek, N.: Evaluation of thunderstorm indices from ECMWF analyses, lightning data and severe storm reports, *Atmospheric Research*, 93, 381 – 396, <https://doi.org/https://doi.org/10.1016/j.atmosres.2008.11.005>, 2009.
- Kunz, M.: The skill of convective parameters and indices to predict isolated and severe thunderstorms, *Natural Hazards and Earth System
600 Sciences*, 7, 327–342, <https://doi.org/10.5194/nhess-7-327-2007>, 2007.
- Lee, J.: Tornado Proximity Soundings from the NCEP/NCAR Reanalysis Data, Ph.D. thesis, University of Oklahoma, 2002.
- Letkewicz, C. E. and Parker, M. D.: Forecasting the Maintenance of Mesoscale Convective Systems Crossing the Appalachian Mountains, *Weather and Forecasting*, 25, 1179–1195, <https://doi.org/10.1175/2010WAF2222379.1>, 2010.
- López, L., Marcos, J. L., Sánchez, J. L., Castro, A., and Fraile, R.: CAPE values and hailstorms on northwestern Spain, *Atmospheric
605 Research*, 56, 147 – 160, [https://doi.org/https://doi.org/10.1016/S0169-8095\(00\)00095-8](https://doi.org/https://doi.org/10.1016/S0169-8095(00)00095-8), 2001.
- Lucas, C., Zipser, E. J., and LeMone, M. A.: Convective Available Potential Energy in the Environment of Oceanic and Continental Clouds: Correction and Comments, *Journal of the Atmospheric Sciences*, 51, 3829–3830, [https://doi.org/10.1175/1520-0469\(1994\)051<3829:CAPEIT>2.0.CO;2](https://doi.org/10.1175/1520-0469(1994)051<3829:CAPEIT>2.0.CO;2), 1994.
- Marsh, P. T., Brooks, H. E., and Karoly, D. J.: Preliminary investigation into the severe thunderstorm environment of Europe simulated by the Community Climate System Model 3, *Atmospheric Research*, 93, 607 – 618, <https://doi.org/https://doi.org/10.1016/j.atmosres.2008.09.014>, 2009.
- McNulty, R. P.: Severe and Convective Weather: A Central Region Forecasting Challenge, *Weather and Forecasting*, 10, 187–202, [https://doi.org/10.1175/1520-0434\(1995\)010<0187:SACWAC>2.0.CO;2](https://doi.org/10.1175/1520-0434(1995)010<0187:SACWAC>2.0.CO;2), 1995.
- Miller, R. C.: Notes on analysis and severe-storm forecasting procedures of the Air Force Global Weather Central, vol. 200, AWS Technical
615 Report, 1975.
- Mohr, S., Kunz, M., and Geyer, B.: Hail potential in Europe based on a regional climate model hindcast, *Geophysical Research Letters*, 42, 10,904–10,912, <https://doi.org/10.1002/2015GL067118>, 2015.

- Molina, D. S., Fernández-González, S., González, J. C. S., and Oliver, A.: Analysis of sounding derived parameters and application to severe weather events in the Canary Islands, *Atmospheric Research*, 237, 104865, 620 <https://doi.org/https://doi.org/10.1016/j.atmosres.2020.104865>, 2020.
- Moncrieff, M. W.: A theory of organized steady convection and its transport properties, *Quarterly Journal of the Royal Meteorological Society*, 107, 29–50, <https://doi.org/10.1002/qj.49710745103>, 1981.
- Nakanishi, M. and Niino, H.: An Improved Mellor–Yamada Level-3 Model: Its Numerical Stability and Application to a Regional Prediction of Advection Fog, *Boundary-Layer Meteorology*, 119, 397–407, <https://doi.org/10.1007/s10546-005-9030-8>, 2006.
- 625 Owens, R. G. and Hewson, T.: ECMWF Forecast User Guide, ECMWF, <https://doi.org/10.21957/m1cs7h>, 2018.
- Parrish, D. F. and Derber, J. C.: The National Meteorological Center’s Spectral Statistical-Interpolation Analysis System, *Monthly Weather Review*, 120, 1747–1763, [https://doi.org/10.1175/1520-0493\(1992\)120<1747:TNMCSS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2), 1992.
- Piper, D. and Kunz, M.: Spatiotemporal variability of lightning activity in Europe and the relation to the North Atlantic Oscillation teleconnection pattern, *Natural Hazards and Earth System Sciences*, 17, 1319–1336, <https://doi.org/10.5194/nhess-17-1319-2017>, 2017.
- 630 Púčik, T., Groenemeijer, P., Rýva, D., and Kolář, M.: Proximity Soundings of Severe and Nonsevere Thunderstorms in Central Europe, *Monthly Weather Review*, 143, 4805–4821, <https://doi.org/10.1175/MWR-D-15-0104.1>, 2015.
- R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>, 2018.
- Rädler, A. T., Groenemeijer, P., Faust, E., and Sausen, R.: Detecting Severe Weather Trends Using an Additive Regressive Convective Hazard 635 Model (AR-CHaMo), *Journal of Applied Meteorology and Climatology*, 57, 569–587, <https://doi.org/10.1175/JAMC-D-17-0132.1>, 2018.
- Rädler, A. T., Groenemeijer, P. H., Faust, E., Sausen, R., and Púčik, T.: Frequency of severe thunderstorms across Europe expected to increase in the 21st century due to rising instability, *Npj Climate and Atmospheric Science*, 2, 30, <https://doi.org/10.1038/s41612-019-0083-7>, 2019.
- Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., and Schlax, M. G.: Daily High-Resolution-Blended Analyses for Sea 640 Surface Temperature, *Journal of Climate*, 20, 5473–5496, <https://doi.org/10.1175/2007JCLI1824.1>, 2007.
- Riemann-Campe, K., Fraedrich, K., and Lunkeit, F.: Global climatology of Convective Available Potential Energy (CAPE) and Convective Inhibition (CIN) in ERA-40 reanalysis, *Atmospheric Research*, 93, 534 – 545, <https://doi.org/https://doi.org/10.1016/j.atmosres.2008.09.037>, 2009.
- Rodríguez-Puebla, C., Encinas, A. H., Nieto, S., and Garmendia, J.: Spatial and temporal patterns of annual precipitation 645 variability over the Iberian Peninsula, *International Journal of Climatology*, 18, 299–316, [https://doi.org/10.1002/\(SICI\)1097-0088\(19980315\)18:3<299::AID-JOC247>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1097-0088(19980315)18:3<299::AID-JOC247>3.0.CO;2-L), 1998.
- Romero, R., Ramis, C., and Guijarro, J.: Daily rainfall patterns in the Spanish Mediterranean area: an objective classification, *International Journal of Climatology*, 19, 95–112, [https://doi.org/10.1002/\(SICI\)1097-0088\(199901\)19:1<95::AID-JOC344>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1097-0088(199901)19:1<95::AID-JOC344>3.0.CO;2-S), 1999.
- Romero, R., Gayà, M., and Doswell, C. A.: European climatology of severe convective storm environmental parameters: A test for significant 650 tornado events, *Atmospheric Research*, 83, 389 – 404, <https://doi.org/https://doi.org/10.1016/j.atmosres.2005.06.011>, 2007.
- Rummukainen, M.: State-of-the-art with regional climate models, *Wiley Interdisciplinary Reviews: Climate Change*, 1, 82–96, <https://doi.org/10.1002/wcc.8>, <https://onlinelibrary.wiley.com/doi/abs/10.1002/wcc.8>, 2010.
- Sáenz, J., González-Rojí, S. J., Carreno-Madinabeitia, S., and Ibarra-Berastegi, G.: Analysis of atmospheric thermodynamics using the R package aiRthermo, *Computers & Geosciences*, 122, 113 – 119, <https://doi.org/https://doi.org/10.1016/j.cageo.2018.10.007>, 2019.
- 655 Showalter, A. K.: A Stability Index for Thunderstorm Forecasting, *Bulletin of the American Meteorological Society*, 34, 250–252, 1953.

- Siedlecki, M.: Selected instability indices in Europe, *Theoretical and Applied Climatology*, 96, 85–94, <https://doi.org/10.1007/s00704-008-0034-4>, 2009.
- Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W., and Bronaugh, D.: Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate, *Journal of Geophysical Research: Atmospheres*, 118, 1716–1733, <https://doi.org/10.1002/jgrd.50203>, 2013.
- 660 Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W., and Powers, J. G.: A Description of the Advanced Research WRF Version 3, NCAR Technical Note NCAR/TN-475+STR, <https://doi.org/10.5065/D68S4MVH>, 2008.
- Taszarek, M., Brooks, H. E., and Czernecki, B.: Sounding-Derived Parameters Associated with Convective Hazards in Europe, *Monthly Weather Review*, 145, 1511–1528, <https://doi.org/10.1175/MWR-D-16-0384.1>, 2017.
- 665 Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, *Journal of Geophysical Research: Atmospheres*, 106, 7183–7192, 2001.
- Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M., Mitchell, K., Ek, M., Gayno, G., Wegiel, J., and Cuenca, R.: Implementation and verification of the unified NOAH land surface model in the WRF model, in: 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction, vol. 1115, 2004.
- 670 Tiedtke, M.: Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models, *Monthly Weather Review*, 117, 1779–1800, [https://doi.org/10.1175/1520-0493\(1989\)117<1779:ACMFSF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2), 1989.
- Tullot, I. F.: *Climatología de España y Portugal*, vol. 76, Universidad de Salamanca, 2000.
- Ulazia, A., Sáenz, J., Ibarra-Berastegui, G., González-Rojí, S. J., and Carreno-Madinabeitia, S.: Using 3DVAR data assimilation to measure offshore wind energy potential at different turbine heights in the West Mediterranean, *Applied Energy*, 208, 1232 – 1245, <https://doi.org/https://doi.org/10.1016/j.apenergy.2017.09.030>, 2017.
- 675 Ulazia, A., Ibarra-Berastegi, G., Sáenz, J., Carreno-Madinabeitia, S., and González-Rojí, S. J.: Seasonal Correction of Offshore Wind Energy Potential due to Air Density: Case of the Iberian Peninsula, *Sustainability*, 11, <https://doi.org/10.3390/su11133648>, 2019.
- van Delden, A.: The synoptic setting of thunderstorms in western Europe, *Atmospheric Research*, 56, 89 – 110, [https://doi.org/https://doi.org/10.1016/S0169-8095\(00\)00092-2](https://doi.org/https://doi.org/10.1016/S0169-8095(00)00092-2), 2001.
- 680 Viceto, C., Marta-Almeida, M., and Rocha, A.: Future climate change of stability indices for the Iberian Peninsula, *International Journal of Climatology*, 37, 4390–4408, <https://doi.org/10.1002/joc.5094>, 2017.
- Virts, K. S., Wallace, J. M., Hutchins, M. L., and Holzworth, R. H.: Highlights of a New Ground-Based, Hourly Global Lightning Climatology, *Bulletin of the American Meteorological Society*, 94, 1381–1391, <https://doi.org/10.1175/BAMS-D-12-00082.1>, 2013.
- 685 Wilks, D. S.: *Statistical Methods in the Atmospheric Sciences*, vol. 100, Academic Press, 2011.
- Xu, G., Xi, B., Zhang, W., Cui, C., Dong, X., Liu, Y., and Yan, G.: Comparison of atmospheric profiles between microwave radiometer retrievals and radiosonde soundings, *Journal of Geophysical Research: Atmospheres*, 120, 10,313–10,323, <https://doi.org/10.1002/2015JD023438>, 2015.
- Ye, B., Del Genio, A. D., and Lo, K. K.-W.: CAPE Variations in the Current Climate and in a Climate Change, *Journal of Climate*, 11, 1997–2015, <https://doi.org/10.1175/1520-0442-11.8.1997>, 1998.
- 690 Zhang, C., Wang, Y., and Hamilton, K.: Improved Representation of Boundary Layer Clouds over the Southeast Pacific in ARW-WRF Using a Modified Tiedtke Cumulus Parameterization Scheme, *Monthly Weather Review*, 139, 3489–3513, <https://doi.org/10.1175/MWR-D-10-05091.1>, 2011.

Zheng, D., Van Der Velde, R., Su, Z., Wen, J., and Wang, X.: Assessment of Noah land surface model with various runoff parameterizations over a Tibetan river, *Journal of Geophysical Research: Atmospheres*, 122, 1488–1504, <https://doi.org/10.1002/2016JD025572>, 2017.