

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

by Santos J. González-Rojí and co-authors submitted to HESS

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.

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.

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, what 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?
- 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?
- 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.
- 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 mechanism 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.
- 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).
- 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.
- 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.

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

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.
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.
3. When describing the general convective situation over the IP / over Europe, you should consider also more recent literature.
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.
5. L1 (see major point above): Instability does not trigger convection.
6. L2 (also L29-30): CAPE/CIN are measures of the energy and not instability indices.
7. Shorten the abstract and focus on the essentials.
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.
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).
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).
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.
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.
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.

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”.
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).
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.
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.)).
18. L57: Explain how the seasonality of precipitation is determined by topography?
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).
20. L72-73: These are very low values. Other studies (e.g., Kunz, 2007, NHESS; Pucik et al., 2015 (MWR), Tazsarek et al., 2017 (MWR)) found much higher CAPE values (also for different version of the mixed-layer CAPE). This should at least be mentioned.
21. Paragraph 57-75: Separate between precipitation and convective environment (CAPE).
22. L76-80: I do not see how climate change is related to this work. I propose to delete this paragraph.
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?
24. L110: Give some more details on the levels: spacing, highest level, which ones are used to compute CAPE/CIN.
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.
26. Section 2.2: Why haven’t you considered IGRA sounding data?
27. L135: For readers outside of Europe it would be helpful to include here also local times (approximately).
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.
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
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 θ_e).
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 $\sim \frac{1}{2}$ hour to reach the LNB.
32. L169: What is “an isobaric precooling” and why was it applied?
33. L172: As TT relies on temperature differences, the unit ($^{\circ}\text{C}$, K) does not matter.

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).
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?
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.
37. Sect. 2.3.2, last sentence: please delete the statement about precipitation (cf. major point 4).
38. L188: "by independent locations": Independent in which sense? The locations are not independent right now.
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.
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?
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?
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?
43. L231: "...small differences in initial conditions..."; can you be more specific here (also with regard to TT, as already alluded above)?
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?
45. L250: Could you be more specific?
46. L254: "N tends to overestimate the variability in every season and for most of the stations..." Why?
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.
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?).
49. L261-262: Again I miss a reference that shows not only total precipitation, but a classification among the types (stratiform/convective).
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?
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?

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.
53. L305: "...lowest values are observed near the coastal valleys..." why?
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?
55. L326 and following: See major comment 5 and minor 52.
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.

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. L1123: 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..."