Review of "Spatial and temporal variability of clouds and precipitation over Germany: multiscale simulations across the gray zone" by Barthlott and Hoose

This paper uses 24-hour simulations for 6 case studies run using the COSMO model at various horizontal grid-spacings (from 250m to 2.8km). An intercomparison of the rainfall, wind convergence, vertical velocity and 2m temperature and humidity characteristics of the different model resolutions is performed and the simulations are evaluated against rainfall observations from radar. The abstract suggests that the paper would be an interesting model resolution inter-comparison using well observed cases from a major field campaign. I was disappointed to find that in fact no observations are used in the study apart from rain radar which presumably is a product available over Germany for all recent periods rather than a special addition for the field campaign. The paper is well written but the analysis is predominately qualitative and there is no evaluation or discussion of clouds at all in the paper, even though the word cloud appears in the title(!). The study does demonstrate qualitatively that finer detail is gained in aspects such as gravity waves and convergence when going to fine resolution (especially Figures 5-7 and 11) but other than this I didn't learn anything new from the results. In particular, I felt that Figures 12-15 didn't really add much to the analysis. Given the significant effort that would have been put into running these cases at such high horizontal resolution, I would have expected a more quantitative and thoughtful analysis and comparison with the available field campaign observations. For this reason I have to recommend that this paper is rejected but could be subject to reconsideration once the analysis has been improved. I expand on some of these concerns below and outline additional major and minor comments.

Further major comments

P17136, L15-19 and elsewhere: You find that the model simulations behave quite similarly in the cases with stronger synoptic forcing. This is likely to be because the lateral boundary conditions (which all come from the 2.8km model) have a significant impact on what happens within the domain. This is quite well known from other studies. I would make this point somewhere in the conclusions.

The labelling of the model simulations varies throughout the paper and is confusing. It doesn't explicitly say anywhere that the 2.8km simulation is the reference run. I assume this is the case? Also, the simulations are given names in Table 1 but these are not used throughout the paper, if at all in the text. I suggest only using these names in all the text and figures and changing C2.8 to C2.8R to indicate it is the reference simulation. Label the 1km simulations C11D and C13D so it is clear which has 1D and which has 3D turbulence. Additionally, in some sections of the text the cases are referred to by date and in others by IOP number. I suggest using the dates at all points in the text/figures.

Figure 2 is a bit unnecessary. I suggest removing it and just explaining in words what you did with the orography. L25-29 on P 17143 could be removed. L3-8 on P17144 could also be removed as it is repetition.

P17147, L8-9 how is the maximum convergence computed? Is it simply the maximum grid-box value at 16:00 UTC? Wouldn't something like the 95th percentile be a better value to quote here?

I strongly suggest combining Figures 4 and 8 i.e. add the observations to Figure 8 and make it Figure 4 because it is very hard to flick between the two figures to assess model performance.

The evaluation of model rainfall is very qualitative (Figures 8 and 9). I think computing pdfs of rainfall, similar to either Fig 4 of Kendon et al. (2012) or Fig 2 of Holloway et al. (2012) would allow a better comparison of rainfall rates.

Kendon et al. 2012, Realism of Rainfall in a Very High-Resolution Regional Climate Model, J. Climate, 25, 5791-5806.

Holloway CE, Woolnough SJ, Lister GMS. 2012. Precipitation distributions for explicit versus parametrized convection in a large-domain high-resolution tropical case study. Q. J. R. Meteorol. Soc. DOI:10.1002/qj.1903.

Figures 5 to 7. I like the way you have demonstrated the change in detail with increasing model resolution. I was wondering if there was a difference between the 1km simulations with the 1D and 3D turbulence schemes? If so add the panels to the figures, if not state this in the text.

Figure 6. The band of rainfall that spans SW-NE from Dortmund (shown in the top row of Fig 8) sits almost exactly in the white region of Figure 6, i.e. where 10m wind convergence is very weak. Convection is usually associated with convergence. Is the convergence reduced after the onset of the rainfall and are the winds divergent at 16:00 UTC due to downdrafts and cold pool outflows? The low-level convergence is plotted at 16:00 UTC and Figure 10a suggests that the rainfall began in the morning.

Figures 9 and 10. There are large differences between the rainfall amounts measured by the radar and those simulated by the models. Radar-derived rainfall products have some uncertainty and an estimate of this should preferably be added to Figure 10. At the very least studies comparing radar rainfall estimates with rain gauges should be referenced to get an idea of how accurate the observations may be.

P17150, L27-30 “The results of our simulations do not show an systematic under or over estimation of the radar-derived precipitation amount”. I don’t agree with this statement because you don’t know if your simulated values are correct or not. It is feasible that the radar consistently overestimates rainfall but the simulations under estimate for some cases and over estimate for others. Deleting this sentence and attending to the previous comment will solve this issue.

As mentioned in the first paragraph of this review, the case studies are labelled as Intensive Observation Periods, which suggests a significant number of observations exist on these days. My understanding is that the KITcube has a whole host of relevant measurements. Why aren’t these and perhaps other observations (e.g. radiosondes to get estimates of CAPE etc) used in this study? This would hugely improve the quality of the study and allow more quantitative model evaluation.

L17155, P5-7 “The dominant value of all runs always has a negative sign” (Figure 11). It looks to me like the dominant value (i.e. the peak of the line) is zero for most/all the simulations.

Figures 12-15 as stated in the first paragraph of the review I didn't really see the point of this whole section of analysis. The way the data is presented is quite hard to follow and I don't feel I learnt anything useful from it.

P17164, L18-21 Why exactly do you recommend the simulation with 1D turbulence over the one with 3D turbulence, when they produce very similar results (for your cases)? This is not justified in your analysis. Is it because the performance is similar but 3D is computationally more expensive? This relates to my previous comment about showing 1km with 3D turbulence in Figures 5-7.

Minor issues

P17136,L8 "COSMO model to real weather" -> "COSMO model to represent real weather"

P171136,L14 "rain intensities **may** vary with resolution, leading to differences in the total rain amount of up to +48%" So do they? Weak statements such as this shouldn't be included in an abstract.

P17141,L10 "In vertical direction" -> "In the vertical direction"

P17141,L16 "for entire Europe" -> "for all of Europe"

P17141,L21 "allows to switch off the parameterization of deep convection" -> "allows the parameterization of deep convection to be switched off"

P17142,L21-24 Further information about the model simulations (if they have parameterised shallow convection (or not), the time step etc) should be added to Table 1.

P17146, L16-19 You state that 1km grid-spacing should be sufficient for capturing gravity waves. This is true for this particular case but not necessarily for other cases and locations that are not studied in this paper. You should add this caveat here.

P17147, L2-4 What is the % for the 2.8km simulation?

P17148,L1 The use of convergence here is confusing because you are talking about the models converging to a single solution, rather than the low-level convergence as discussed in the previous paragraphs. Please revise this.

Figures 13 and 14. The deviation of C2.8 from C2.8 should be zero, yet this appears as a red/pink colour on the plots. Surely a zero deviation should represented by the white segment of the colour scale on these plots?

All Figures – label the panels (a), (b), (c) and so on.

P17164, L26 "The large jumps in the dominant values **could** be attributed to the existence of bimodal distributions" so are they?