We thank the reviewers for reading the manuscript and providing detailed comments. We have carefully considered all comments and changed the manuscript accordingly. Please find below our responses in italic and blue.

Anonymous Referee #1

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 intercomparison 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.

We appreciate the reviewer's constructive arguments to give us the possibility to be reconsidered for publication. The first major criticism is the fact that no HDCP2 observations were used to evaluate the model. However, this was not the goal of this study. First, we want to describe the resolution effects on the variability of precipitation and convection-related parameters and how this variability changes with model resolution. Second, the applicability of the COSMO model to represent real weather events at such high resolutions is tested. The authors are not aware of any COSMO results in the peer-reviewed literature at such high resolutions (apart from idealized studies), which emphasizes the contribution of this work. At the time where most of the actual work was done, not much quality-controlled IOP data was available from the HOPE field campaign. Therefore, we focused on the resolution sensitivity which, in our opinion, is a fruitful topic on its own and left the comparison with observations to a future work. The incorporation of such an evaluation would also blast the length of the present manuscript.

The second main concern of the reviewer (also of reviewer #2) is the missing discussion of clouds in the manuscript. We have therefore introduced a new section 3.5 with the analysis of the liquid water path about its temporal evolution and spatial coverage. Additionally, a new figure with vertical profiles of cloud condensate was added to the last section on grid spacing effects for the summertime case. We also like to point out that the parameters analyzed in Figs. 12-15 are strongly related to the cloud development.

The reviewer states that Figures 12-15 did not add much to the analysis. In our opinion, this analysis is important to describe the variability of these parameters with model resolution. However, we left out the section with the analysis of the humidity index HI which provides only little additional information. The old Figure 14 showing the dominant value of this index has been deleted as well.

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.

We agree with the reviewer. However, we already had a sentence in the conclusions about that fact: „We therefore conclude that the synoptic forcing plays a larger role for the HOPE cases as for the summertime case, where no synoptic-scale upward forcing is present and that this larger role somehow limits the possible effects of a higher grid resolution. Similar to findings of Talbot et al. (2012), the data needed to force the individual domains and to initialize surface parameters have the strongest influence on the results.“ We believe that this statement is sufficient.

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.

Yes, the 2.8 km run is the reference run. We added „reference run“ in Table 1 and also in the section 2.1 describing the numerical model. However, we did not change C2.8 to C2.8R because we think the new text is sufficient. We did label the 1 km simulations „C1-1D“ and „C1-3D“ to better distinguish between 1D and 3D turbulence scheme. This was done in the entire text and in all figures.

In the text, we now only use the date and not the IOP number. In some figures, however, both informations are given to better refer to the individual IOPs.

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.

We removed this figure as suggested and adapted the text accordingly. However, we kept the text from L25-29 as it gives the some information from the removed figure. We deleted L3-8 on P17144 as suggested.

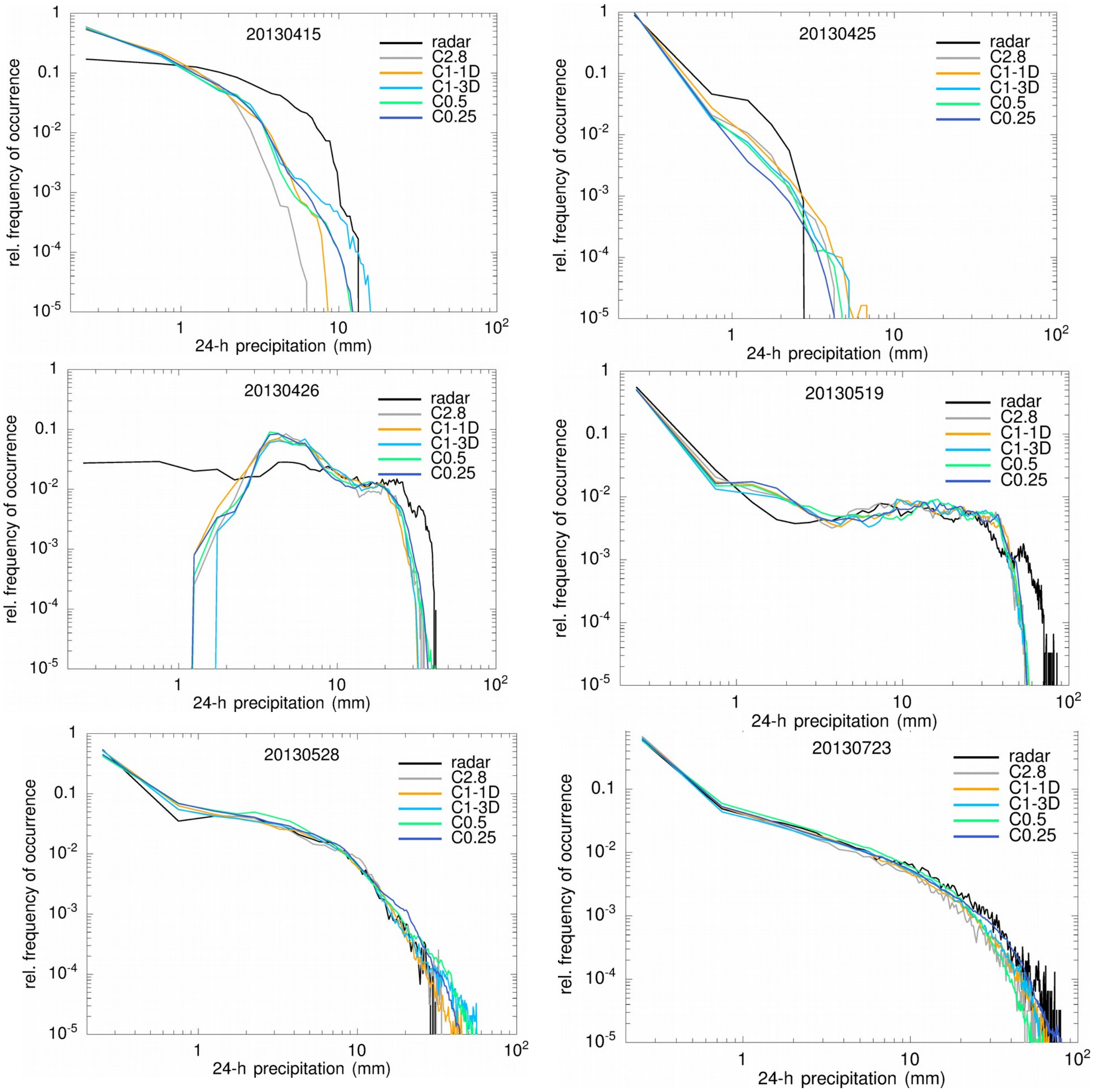
P17147, L8-9 how is the maximum convergence computed? Is it simply the maximum grid-box value at 16:00 UTC? Wouldn't something like the 95 th percentile be a better value to quote here? I strongly suggest combining Figures 4 and 8 i.e. add the observations to Figure 8 and make it Figure 4 because it is very hard to flick between the two figures to assess model performance.

It was simply the maximum convergence of all grid points inside the area of investigation. We added the 95th percentile and put those numbers in the new Table 3 for better readability.

We agree with the reviewer that it is not ideal to flick between Figures 4 and 8. The reasons why we left it that way are the following ones: (i) The days under investigation are described in section 2.2. We think it is important to have the rainfall information already here. (ii) The figures cannot be *combined because otherwise, the individual subfigures would become too small. One still could zoom in in the electronic version, but you would not be able to see all details in a printed version.*

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.

We believe that the information given in Figure 9 is sufficient for our purpose: one can see the median, the maximum, and the 25th/75th percentiles of the rain distributions. To include the PDFs of all days analyzed would require too many figures for comparatively little additional information. Instead, those PDFs are presented here in our reply. For the majority of days analyzed, the PDFs of simulated and radar-derived rain are similar. Disregarding the days with no or very little rain (24+25 April), the most pronounced difference can be found on 26 April where the minimum rain amount in the simulations is always higher than in the observations where also rain-free areas are present.

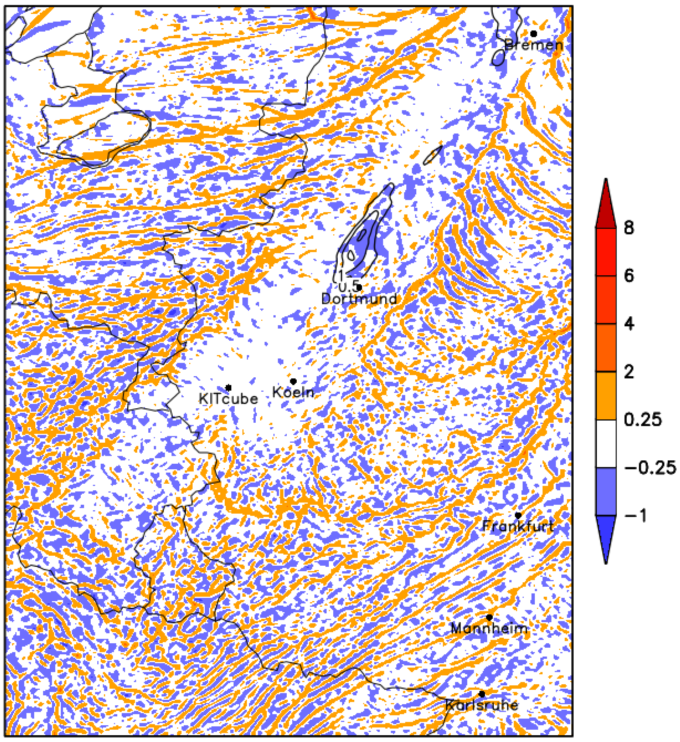


Figures 5 to 7. I like the way you have demonstrated the change in detail with increasing model resolution. I was wondering if there was a difference between the 1km simulations with the 1D and 3D turbulence schemes? If so add the panels to the figures, if not state this in the text.

There was no significant difference for these features in both 1 km simulations. We added a remark at the beginning of subsection 3.1.

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.

The reviewer's assumption is correct: The winds in that region are rather divergent due to weak rainfall. The figure on the right shows convergence in red and divergence in blue with black isolines of 30 min rain rate for the same time as in Figure 6 in the manuscript (run C1-3D). The white area in Figure 6 shows primarily divergent areas where precipitation is falling and otherwise only weak convergence and divergence. Therefore, when we display only the convergent contributions as in Figure 6, this area remains more or less white.



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.

There is already one statement with a former study in the manuscript in section 3.3 "Although radar is not an instrument measuring precipitation in a quantitative sense (see e. g. Rossa et al., 2005)...".

Rossa et al.: COST 717 Final Report, "Use of Radar Observations in Hydrological and NWP Models"

To add an estimate of the uncertainty for the radar network used in our study cannot easily be done. Since we focus on the spatial distribution and temporal evolution, we believe this is sufficient that way.

P17150, L27-30 “The results of our simulations do not show a 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.

We agree with the reviewer and deleted that sentence.

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.

Please see also our reply to the referee's first comment.

The main reason why we did not use any HOPE observations is that a systematic model evaluation was not the goal of our study. Our focus was (i) the resolution dependance of several convection-related parameters and (ii) the testing of the COSMO model for real weather events at very high resolutions. The incorporation of such an evaluation would blast the length of the already long manuscript. Therefore, the systematic comparison with observation is left to future work.

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.

The dominant value is close to zero, but has a negative sign.

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.

Please see also our reply to the referee's first comment.

In our opinion, this analysis is important to describe the variability of these parameters with model resolution. The parameters analyzed are also related to clouds and the way the results are presented is new (to the author's knowledge). We think that this concise representation may also be helpful for future studies. However, we deleted parts of the text (concerning the humidity index HI) and Figure 14 which are not essential for this study.

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.

As the 3D scheme is performing similarly as the 1D scheme (regarding precipitation amount and timing), we recommend the 1D scheme which is indeed computational slightly less expensive. We rephrased the text slightly and also included some findings about the liquid water path and the cloud condensate.

Minor issues

P17136,L8 “COSMO model to real weather” -> “COSMO model to represent real weather”

done

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.

We replaced „may vary“ with „vary“.

P17141,L10 “In vertical direction” -> “In the vertical direction”

done

P17141,L16 “for entire Europe” -> “for all of Europe”

done

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

done

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.

We included this information in Table 1 as suggested.

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.

We refined this sentence accordingly.

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

The percentage cover of grid points with wind convergence in the 2.8 km run is 44%.

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.

We rephrased that sentence.

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?

Yes, it should be a white segment since the deviation is zero. In our figures, it is correctly represented in white, we checked it with a picture editing program by determining the RGB values.

We get the same values as the white surrounding.

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

done

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

Yes, they are. We changed „could be“ to „were“.

Anonymous Referee #2

In their manuscript Barthlott and Hoose aim to present a study on the dependence of precipitation and convection patterns on the grid scaling of the COSMO model. The main messages of the publication are that convective processes are increasingly better resolved with increasing resolution. This affects the formation of precipitation, even though it was not found to actually improve the quantitative precipitation forecast (except for some indication in the summertime case). The paper is mainly of qualitative nature. The discussion and presentation of the results are done adequately. Actual quantitative information or a clear definition of pro's and con's of increasing/changing the horizontal resolution are however not given which reduces somewhat the scientific value of the publication. I consider the scope of the paper to be a well readable documentation of scaling effects on precipitation processes which is in a rather publicable state.

We thank the reviewer for theses comments. We believe that we already gave some definitions of pro's (low-level convergence, PBL thermals, gravity waves) and con's (rain amount and timing similar in situations with large synoptic forcing).

Two major issues however still remain. Since clouds are not treated in the manuscript, the title and content of the manuscript need to be modified. If the title cannot be changed, the paper needs to be withdrawn and resubmitted or the manuscript needs to be extended. Taking into account the second comment would improve the scientific value of the manuscript considerably. See below for details.

We agree with the reviewer and changed the content of the manuscript accordingly and left the title unchanged. We introduced a new section 3.5 with the analysis of the liquid water path about its temporal evolution and spatial coverage. Additionally, a new figure with vertical profiles of cloud condensate was added to the last section on grid spacing effects for the summertime case. We also like to point out that the parameters analyzed in Figs. 12-15 are strongly related to the cloud development.

Major comments:

1) The term 'clouds' in the manuscript title is misleading because no information on the presence of clouds is given at all. This also affects the content of the paper (p17136 lines 1 and 19; p 17144, line 4). Consequently, if the authors want to stick to the presence of the term 'clouds' in the title, they actually are required to present information on cloud properties. There are probably quite some indices available to do so, but validation will be more difficult because measurements of the vertical structure of clouds are rare. Even though they are available with rather high density within the HOPE domain (from cloudnet).

Please also see our reply to the previous comment.

2) The reason why I consider the paper to be (just) "a well readable documentation of scaling effects on precipitation processes" is the missing of quantitative information on reasons for the observed differences. Precipitation produced by a model is the result of a large number of model processes and parametrizations that interact with each other. By looking just on the precipitation and the convection fields, one cannot resolve the processes taking place in between – which are likely the reason for resulting differences. It is well known that the resolution of a model cannot be increased without checking for the validity of involved parametrizations. Keyword here is the 'effective resolution' of a model. Below that effective resolution turbulent/kinetic processes are not resolved accurately (Skamarock, 2004 (Section 3), Petrik 2012). How is this problem dealt with within the present study? There should be a section added to the manuscript to discuss the discrepancy between actual model resolution and effective model resolution. Is it possible to show

the energy spectrum for the different resolutions? If there is a dependence of the spectra on the model resolution, how will this effect the forecast efficiency? E.g., there may be increasing turbulence with increasing model resolution. But does the energy spectrum show the expected 5/3 decay?

We believe that at least some quantitative information on reasons for the observed precipitation differences are already given in the paper. E.g. section 3.4 analyzes the differences in precipitation based on a number of convection-related variables like CAPE and CIN, low-level wind convergence and convergence-induced lifting of air parcels.

The checking of the validity of the involved parameterizations and the effective resolution is a very good hint for an additional subsection. As requested by the reviewer, we computed kinetic energy spectra for the different model resolutions and included a new subsection 3.2 with an additional figure. There is also some new text in the abstract and the conclusions.

Minor comments:

1) I would stay with either the term ‘terra incognita’ or ‘gray zone.’ Don’t switch between both in the text. ‘gray zone’ is currently used more often than ‘terra incognita’. Thus staying with gray zone would ease the readability. Introduce ‘terra incognita’ only in the introduction.

Thanks for pointing that out, we adjusted the manuscript accordingly.

2) Section 3.1 only deals with the vertical velocity/convection. The section title can be adjusted accordingly.

We agree with the reviewer that all analyzed processes are about or linked to vertical velocity and convection. However, because low-level wind convergence is just a precursor of vertical movements and gravity waves are not necessarily linked to convection, we changed the title to „Benefits of high-resolution modeling for gravity waves, low-level wind convergence, and PBL thermals“.

3) Acknowledgements: Shouldn’t HOPE/HDCP2 be acknowledged, too?

We did not receive any direct funding for this work, but we included HOPE and HDCP2 in the Acknowledgements.

References:

William C. Skamarock, 2004: Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra. Mon. Wea. Rev., 132, 3019–3032. doi:http://dx.doi.org/10.1175/MWR2830.1

Ronny Petrik, 2012 : Physical validation and bracket-based dynamical cores for mesoscale NWP models. Phd thesis in“Berichte zur Erdsystemforschung”, 121, ISSN 1614-1199. Available at http://www.mpimet.mpg.de/wissenschaft/publikationen/berichte-erdsystemforschung.html