2nd Review of "Divergent convective outflow in ICON deep convection-permitting and parameterised deep convection simulations", by Edward Groot, Patrick Kuntze, Annette Miltenberger, and Holger Tost.

March 5, 2024

Many thanks for the updated manuscript. The correction to figure 6 has addressed my biggest concern; the divergence profiles plotted for PER just had an erronious factor of air density; there is no sign that anything else has gone wrong.

I am happy to accept the paper, pending some minor corrections discussed below.

My requests for a more "like for like" comparison between the ICON-PER and ICON-PAR simulations, by using identical analysis methods for both, have not been addressed. But I am satisfied that making these changes to the method would be very time-consuming / require rerunning computationally-expensive data-processing, and would be unlikely to significantly affect the results. So I am happy to accept the article without this change.

I thank the author for acknowledging / clarifying that the study did not find sensitivity of the precipitation - divergence relationship to the choice of parameterised versus explicit convection; the sensitivity found appears to be dominated by model-resolution alone. This point just ought to be clarified in a couple other places in the paper (I give some suggestions below).

New reviewer replies to the authors' response are given below in **blue**, following the original numbered list of queries. Below we follow the authors' colour-scheme by quoting the original review queries in **red**, and the authors' responses to these in **black**. Where line numbers are given, these now refer to the authors' track-changes file (not the original or updated manuscript).

1. The request to clarify that the results apply for the transient response of upper-level divergence, not the eventual balanced response.

   Many thanks for the additional discussion L74-83, and for clarifying in various places that the transient response is studied here. This nicely addresses the query.

2. The query about dependence of upper-level divergence on the vertical structure of the latent heating profile, not just the vertically-integrated latent heating.

   ...if we assume that the warm air mass is relatively homogeneous (to the east of the frontal zone, which is over the French-German border region) and that the convection takes place on the flanks of this warm air mass, there may not be sufficient variation in stability profiles among all the simulations to cover this aspect of vertical variability with our case study (studying the convective systems in the warm air mass over Germany, directly east of the frontal zone).

   My argument does not depend on any horizontal variation in the stability profiles, or variation in stability between different models. Rather it is variation in the vertical structure of the heating profile, and how this correlates with *vertical* variation of environment stability. e.g. consider two different convective storms occuring in the same environment and static stability profile, and suppose (as is often observed) that the static stability N^2 is greater in the lower free troposphere than it is in the upper troposphere. Now, suppose storm 1 is a relatively young cell, so that most of the latent heating occurs in the lower troposphere, whereas storm 2 is a mature system with a substantial stratiform anvil region generating more latent heating in the upper troposphere. Even if the storms are producing equal precipitation rates, storm 2 will produce stronger upper-level divergent outflow, just because the heating is concentrated at upper-levels where N^2 is less, so that stronger ascent is needed to make the buoyancy removal by vertical advection ($-wN^2$) balance the heating rate ($\frac{g}{T_v}\frac{\partial T_v}{\partial t}_{latent}$). Even under a horizontally homogeneous environment and uniform neutral buoyancy level, storms in different states of organisation may produce differently-shaped vertical profiles of heating rate, which will give different upper-level divergence if N^2 varies with height?

   We have looked into this in our previous work, [7]. There, we induced variation in the heating profiles by manipulating the constant of latent heat release.

   Many thanks for pointing out the interesting experiment in the authors' LES study, where the latent heat of condensation was varied. Note however that the idealised temperature profile used here (from Weisman & Klemp 1982) has arteficially weak variation of N^2 with height (potential temperature increases nearly linearly). So this may not have sampled the sensitivity to vertical structure.

   I don't think the present study requires a detailed investigation into dependence of outflow strength on vertical structure (this might be an interesting topic for future work). But it might be worth breifly mentioning this possibility in the discussion?

3. Why would it not be possible to represent the impact of storm morphology on upper-level divergence patterns in a simulation at 13km resolution with

parameterised convection?

...the truncation scales (which will strongly suppress variability at sub-100 km scales in the ICON-PAR)...

...we do not agree that the ellipses are (at least typically) fitted to precipitation fields at scales that are resolvable in the parameterised convection setup: the truncation of the grid scale is typically at wavelengths of about 6-8 grid cells of the grid spacing [e.g. [8]], which translates to about 100 km for ICON-PAR. Everything smaller than this scale is in the grey zone and partially resolved to unresolved, with a rapid decrease of the resolved fraction, presumably especially at 4-8 times grid spacings. The size of large ellipses grows to about 100 km in ICON-PER, but there are plenty of smaller ellipses in our dataset as well. Therefore, we are a little concerned that the reviewer may just be too optimistic about the ability of numerical models to resolve mesoscale processes at lengths less than 8 times grid spacing...

The reviewer's experience with NWP models is that they routinely do produce both gravity-waves and updrafts / downdrafts very near the grid-scale (excessive single-grid-point updrafts are a common problem!) Whilst a model can only simulate these features with significant numerical error, the poor resolvability does not stop them from happening. The idea that models contain a truncation scale of $\sim$8 grid-lengths below-which all features are strongly damped has come from large eddy simulation, where this has been deliberately imposed by specifying a suitably large mixing-length parameter in the 3D Smagorinsky turbulence scheme (or by using a highly diffusive numerical method for advection). Imposing such an $\sim$8 grid-length truncation scale on an NWP model (with 13km grid-size and parameterised convection) would not be done operationally, as it would degrade the model skill. So I don't agree with the author; features significantly smaller than 100km can and will be simulated by the 13km NWP model, but with diminished accuracy. This is evident in figure 7(c,d); the ensemble standard deviation of divergence in ICON-PAR has a lot of structure on scales of around half a degree ($\sim$50km), and the individual ensemble members will presumably contain smaller-scale structure that gets smoothed-out when looking at the standard deviation over all members?

The results shown in figure 8b show that the divergence-precipitation relationship is not sensitive to the choice of parameterised versus explicit convection at 13km resolution, so the paper's argument that the overly-linear response in PAR is due to the convection parameterisation (e.g. L680, L710) is not supported.

We do agree that the statements in the paper could be interpreted as overstatements of the contrasts, which is indeed suggested by the data points in Figure 8 that compare the 13 km ICON with and without convection parameterisation. We have revised the implied causalities and methodological descriptions along this line...

Many thanks for the changes to the text following this. But there are a few other parts of the paper which still imply that the divergence-precipitation relation depends on the choice of parameterised versus explicit convection, even though no evidence for this has been presented (the sensitivity is dominated by model-resolution alone, as is now stated by the added text at L467).

L750-751: maybe change:

"feedback from deep convection to its surroundings at larger scales is likely not accurately represented with parameterised deep convection, even if the precipitation climatology is well represented: parameterised ensembles seem to be underdispersive in terms of corresponding dynamical variability at a given precipitation rate"

to something like:

"feedback from deep convection to its surroundings at larger scales is likely not accurately represented *at coarse resolution*, even if the precipitation climatology is well represented: ensembles *with 13km grid-size* seem to be underdispersive in terms of corresponding dynamical variability at a given precipitation rate"

Maybe also clarify this in the abstract for consistency, e.g. change L3-4:

"Near-linear response of deep convective outflow strength to net latent heating is found for parameterised convection, ..."

to e.g.

"Near-linear response of deep convective outflow strength to net latent heating is found for 13km grid-spacing with either parameterised or explicit convection, ..."

In addition, to be more specific on the implications for resolving gravity waves, the ICON-PAR grid does not allow to represent gravity wave sources and the full spectrum of relevant gravity waves that the divergence consists of. It may mathematically be considered as a coastline problem in 3D: if we represent an island of 100km2 on a 10 km grid, its coast line would typically be 40 km. However, if we can represent it more accurately at refined grids, the length may deviate strongly.

Many thanks for this insight! While a significant amount of mesoscale structure (in both the updrafts and the radiated gravity-waves) will be present on the 13km grid, it might be the smaller (km-scale) variability *within* the wider mesoscale structures that is key? Even though a 50km-long island can be represented on a 13km grid, the bays and inlets along its coast cannot. Refining the grid to 1km may increase the length of the coastline by an order of magnitude. By analogy, the possibility for gravity-waves radiated from within the "island" to collide with eachother from different directions will vastly increase?

But perhaps clarify at L462-463 that with "little or no information on geometry of the convective systems can be represented", this is really about geometry of smaller-scale structures *within* the convective systems (i.e. km-scale, rather than mesoscale)?

4. Why is low-pass-filtering on a 45km scale applied to ICON-PER (1km) but not ICON-PAR (13km)?

   In retrospective, it might possibly seem a better choice to consider multiple filtering operations, including no filtering, and compare a couple of those options between the two simulation methods as well. On the other hand, this would have created a very comprehensive study and data processing procedure, probably a too comprehensive one for just one manuscript (meaning that two to three manuscripts could be needed to disseminate the results).

   No, my argument was that the analysis would be simpler and the manuscript shorter if the low-pass filter was just omitted (it should make little difference as long as the box-averages consider areas much bigger than the filter-scale).

   Nevertheless, the filtering procedure ought to serve the purpose of creating comparable datasets in simulations of different resolution, rather than diminish comparability (as possibly suggested by the reviewer). This is because the operation should, when applied to ICON at 13 km grid spacing, filter out very little variability: the dynamical spectra of a numerical model drops rapidly at length scales of about 6-8 times grid spacing...

   As discussed earlier, this notion is consistent with LES, but is often not the case in NWP models (especially at coarse resolution with parameterised convection, when they typically do not employ a highly diffusive 3D turbulence scheme to enforce the truncation scale). On the otherhand, it might be that ICON is more diffusive than other NWP models that the reviewer is familiar with. This is critically dependent on the model's advection scheme and any horizontal diffusion used (e.g. as part of a 3D turbulence scheme).

   We have selected a data processing procedure and are confident that it may slightly affect individual samples in mostly random directions (e.g., as the reviewer mentions, by incidently moving some random divergence just to the other side of box boundaries). The assessed conditional statistical signals, however, will not be affected by such processing issues that average out over the lifetime of the systems and, even further, over the full dataset.

   I am happy to take the authors' word that the low-pass filtering does not significantly affect the results anyway. So it is probably not worth rerunning significant parts of the work-flow to pursue this further.

5. Why not define moving boxes in ICON-PAR, to be fully consistent with the analysis of ICON-PER?

   We could have made the choice to further investigate the systems in ICON-PAR. However, this has not been done, as from each system, we have analysed several ensemble members (and physically perturbed simulations) to make sure that the statistical sample size allows for safe conclusions. Furthermore, there are fewer convective systems in these ICON-PAR simulations...

...investigating residual variability and increasing the sample size was clearly not prioritised over the corresponding more detailed analysis of ICON-PER.

This is a reasonable argument; the moving boxes in ICON-PER have been setup manually, repeating this exercise for every ensemble member and sensitivity-test within ICON-PAR would be very laborious, and the effort should be focussed on analysing the more interesting behaviour found in ICON-PER. However, one of the paper's main conclusions is drawn from comparing ICON-PER with ICON-PAR. It would therefore be stronger if more attention was given to ensuring like-for-like comparison and using the same analysis methods when comparing them. e.g. just for the comparison with PAR, the divergence-precip relation in PER could be computed on the same large static boxes as PAR (while still using the smaller moving boxes for the more detailed investigation of relationship to storm structure etc in PER)

But again, at this late stage it may not be worth rerunning the work-flow for a consistency-check which is unlikely to affect the results much.

6. Figure 6: Do the profiles of ensemble-mean box-mean divergence satisfy mass continuity?? In ICON-PER, it looks as though the vertically-integrated lower-troposheric convergence is significantly greater than the upper-tropospheric divergence...

   There is indeed a mistake in Figure 6, where the ICON-PER data display mass divergence data rather than divergence data. Therefore, Figure 6 has been corrected accordingly in the updated manuscript, leading to balanced profiles. This issue has not been found in Figure S5, and Figures 6b and S5 should display similar divergence profiles for ICON-PAR, accounting for the respective pressure axis, which means they are at/very close to mass balance. We thank the reviewer for the close and attentive inspection of the figures, which allowed us to correct this inconsistency in Figure 6!

   Many thanks for investigating and correcting this. It is very reassuring that it was essentially only the units shown on the colour-scale (6a) and x-axis (6b) that were wrong before! This has been corrected by removing an erronious factor of air density from the plotted data, so that they are now consistent with the stated units. This increases my confidence in the paper overall, by removing my worry that something else had gone wrong with the analysis.

7. Why does the analysis of PAR use much bigger boxes than PER?

   The size of the boxes could unfortunately not be chosen smaller than 2-4 degrees in the ICON-PAR; the convective systems inevitably grow larger in this simulation...

...Nevertheless, carefully studying the divergence at upper levels, the convective systems and precipitation rates in ICON-PAR, we have assured that locations with variability in divergence are included in our analysis. This strategy maximises the reliability of the box analysis at the minimum of cost of included remote subsidence (based on the discussion of [7]).

As discussed earlier it would be possible to ensure like-for-like comparison between PAR and PER, by separately computing the precipitation and divergence in PER over the same large static boxes as were used for PAR. But again it is probably not worth rerunning such a large part of the work-flow at this late stage.

I am not overly concerned about the subsidence assumption in the convection parameterisation here. In the reviewer's experience with mass-flux convection schemes, the assumption of "local mass compensation" has little impact, as the model's dynamical core still acts so-as to quickly restore a Weak Temperature Gradient in the horizontal after the parameterised convective heating is applied. This has the effect of immediately undoing the local subsidence imposed by the convection scheme, and redistributing it elsewhere, so that the end result is the same as explicit convection with the equivalent heating profile. The present study's results in figure 8b comparing the divergence response for parameterised versus exlicit convection at 13km grid-size support this.

8. figure 8: The text says that "from here on vertically integrated values (of upper-level mass divergence) are used". But the quantity plotted on the y-axis says it has units kg m^{-3} s^{-1}...

   Essentially, the above approach is equivalent to what has been done for this work. Vertical integration of divergence over the whole layer depth causes downdraft divergence at lower levels to be included (which is partially visible as mean feature towards the end of the time series in Figure 6). By setting a mask this issue is resolved, as the reviewer states. Exactly this has been done. Nevertheless, substantial divergence features from downdrafts do vertically overlap with outflow from shallow convection in ICON-PER, which means that the perfect vertical mask does not exist. Whether the processing leads to a divergence in kilogram per cubic or squared meter depends on whether one concludes by dividing the whole integral by length scale L or not, to obtain the mean value of the integrand. Both are leading to exactly the same result [while the way it is currently plotted, one can also roughly intercompare the local intensity variation of outflows (integrand), which means it is purely a matter of taste, which of the two is favored]. The manuscript has been corrected accordingly.

   Thank you for the edited text at L415, which now clarifies that the quantity shown is the vertical integral divided by outflow depth. Omitting the normalisation by outflow depth would have made the comparison in figure 8 between PER (8a) and PAR (8b) simpler, allowing a slightly more succinct manuscript (the discussion about the correction for differing outflow depths would not be needed). But it does not matter that much.

9. Unclear direction of causality in the correlation between CMT and outflow strength.

   This is a very interesting comment; there is indeed no complete certainty about causality, based on the correlation signal. We agree that both pressure perturbations and momentum/accelerations by turbulent motion can affect the upper tropospheric flow, next to the dominant cause of divergence - the heating in the middle troposphere, which partially arises from pressure perturbations...

   Many thanks for the extensive discussion around the relation between heating, the pressure field, and vertical momentum transport (and the dependence of the "turbulent" transport on model resolution). This made for an interesting and thought-provoking read, and I agree this wider subject is beyond the scope of the present study.

   I am happy with the slight rewording at L612-613, in the synthesis section. Maybe consider also mentioning the uncertain direction of causality in section 6.3, e.g. at L568-569?

10. Minor corrections.

    Many thanks; all the previously listed minor corrections have been addressed. Except:

    L263: "Ellipse fitting and verification are used to quantity the geometry" should be "Ellipse fitting and verification are used to *quantify* the geometry?"

    I also just spotted one more trivial typo:

    L562: "...while Figure Figure 10b shows..."