Dear Reviewer#3,

Thank you very much for your comprehensive review of our manuscript. Please find below our replies to your comments. Note that your comments are written in blue while our replies are black.

Reply to your comments of 18 July (os-2019-44-RC4.pdf)

Major comments

The authors present an interesting numerical experiment for studying the evolution and characteristics of submesoscale processes. It is worth to be published when appropriately revised, especially in regard to the interpretation of results.

(1) My major recommendation is that the manuscript has to be rewritten to state clearly the aim of the study and the methods applied and to interpret the results accordingly. Authors present a numerical experiment where atmospheric forcing was turned off for the model with the highest resolution while the initial and boundary conditions were created with the model with atmospheric forcing. It is unclear what we could learn from such an experiment regarding submesoscale processes. Is the aim to show how the submesoscale processes evolve in case of the sudden vanishing of external forcing and what are their characteristics in such conditions?

See below **(8c), (13), (14), (16), (18), (20), (28), (36)**

Please note as well the new Section 4.3 “Impact of atmospheric forcing“ and the new Fig. 8.

(2) The other concern is related to the description of methods. You state that special care was taken for the preparation of initial and boundary conditions for the R100 model as the nesting ratio of five is challenging. However, you do not provide any arguments on why the cubic spline was used. Did you try other methods or run any sensitivity tests? Are some results (as the revealed false patterns) somehow related to this procedure? No information is given about the atmospheric forcing in HBM. It is not clear how the domain averaged TKE and cumulative averaged TKE are calculated, etc. See the specific comments below.

Action: See below **(5b), (7), (8a,b), (11), (12), (21), (25), (39)**

Specific comments

(3)P1L8: Are the speeds of 100 m day-1 characteristic for submesoscale processes or internal waves?

The horizontal scales of submesoscale processes span the range between 10 m and 10 km and time scales from hours to days. Therefore, internal gravity waves (IGWs) themselves are a submesoscale process besides fronts, eddies, filaments, and frontogenesis. Each of these processes creates strong vertical motions and it makes no sense to sort out the contribution of either process to the amplitude spectrum of the vertical velocity.

Action: none

(4)P1L10-12: Sentences like “The conditions for inertial and symmetric instability are evaluated for the whole domain, and the components of the tendency equation are computed in a subregion.” are not informative in the concluding section of the abstract.

An abstract summarizes the major aspects of the entire paper, including (i) the overall purpose of the study and the research problem, (ii) the design and the methods, and (iii) major findings. The above mentioned sentence belongs clearly to item (ii) and should not be omitted.

Action: none

(5a)P5L9-10: Why HBM daily mean fields were used?

The task of R500 was to provide a realistic mesoscale environment for the 100-m nest. As the shortest time scales of mesoscale processes are on the order of days, the daily averaged fields of HBM were considered to be sufficient for driving R500 at the open boundaries. Moreover, the data volume of the daily averaged fields is around 600 MB for the entire month of June 2016, while the hourly fields would comprise close to 15 GB. That would clearly overschoot the mark.

Action: none

(5b)P5L9-10: What about the atmospheric forcing for HBM?

HBM is run operationally at DMI and BSH, where the BSH run is the backup run in case the DMI run fails. While the DMI run used DMI-HIRLAM forcing at 5 km resolution, COSMO-EU forcing at 7 km is applied with the BSH run. We could easily provide this information in the manuscript, but in our opinion it is not relevant because we used the HBM output *as is*.

Action: none

(6)P6L4-5: What is meant by the term “cycle”? How can the whole first cycle be used as initial conditions?

The HBM daily mean output at CMEMS is provided in NetCDF files which contain all prognostic variables, i.e. the 3D fields T, S, U, V, and the 2D field SSH. For the initialization of R500 and the provision of the open boundary conditions, we downloaded those fields for the entire month of June 2016 in one single file. Hence, this file contained the prognostic variables at 30 time levels on 1, 2, 3, …, 30 June. The data set for each time level is what we call a “cycle“.
The first cycle contains the prognostic fields of 1 June, and that is used for the initialization of R500. From the remaining cycles of 2, 3, 4, …, 30 June we used only the prognostic fields along the open boundaries of the R500 domain.

Action:
In order to avoid any misunderstanding, we renamed “cycle“ to “record“. See

- new manuscript P5L5,31,32

(7)P6L12-13: The use of cubic spline is not justified. Did you run any tests how the interpolation method could influence the results?

In preliminary test runs with linear interpolation, the relative vorticity of jet flows into the R100 domain (i.e. the flow across the open boundaries) looked unrealistic, because the width of cyclonic and anticyclonic shear zones was frequently the same. This is contrary to experience where the width of the cyclonic shear zone is narrower than that of the anticyclonic shear. Using spline interpolation, those jets looked more realistic even though not perfect. The latter applies predominantly to strong jets exceeding 20 cm/s, see new Fig. 7a,b: (i) the relative vorticity is unrealistic in the northwest corner and along the northern boundary of the domain. (ii) the tiny streak with anticyclonic vorticity < -1 at the eastern boundary at 55° 25‘ N is unrealistic.
The above issue is an intrinsic problem of downscaling. We have discussed it with other modelers but a perfect solution is not yet available.

Action: A piece of text was added to Section 2.4. See new ms P6L5-12

(8a)P6L21-23: Atmospheric forcing for R500 and R100 is mentioned here for the first time.What about atmospheric forcing in HBM?

Action: see above **(5b)**

(8b) Do the forcing sources differ between HBM and ROMS?

Yes

Action: none

(8c) Mentioning of atmospheric forcing and interpolation of forcing parameters for R100 is not relevant since you present the results from runs without atmospheric forcing.

Action: As in the revised version, we refer also to an R100 run with atmospheric forcing, the corresponding piece of text in Section 2.4 was modified. See P6L19

(9)P7L1-2: I do not understand this statement about salinity as the ideal parameter. Your atmospheric forcing has a resolution of 6.5 km. What could cause the blurring of the surface signal of temperature in submesoscale range when using such forcing?

It is true that the resolution of the atmospheric forcing is 6.5 km. Hence, the downward shortwave radiation and the infrared back radiation from the clouds are almost (*almost* because of the cubic spline interpolation) identical everywhere in an area of 6.5 km $\times$ 6.5 km. However, this is not true for the other components of the heat budget, i.e. the longwave radiation flux from the ocean, sensible and latent heat fluxes. These quantities depend on the sea surface temperature SST, and the SST in turn is affected by the 3D velocity field and vertical mixing which are subject to submesoscale spatial variability. Therefore, the net surface heat flux varies on the same spatial scales and may blur the submesoscale SST distribution.
In contrast, the surface salinity in the 6.5 km $\times$ 6.5 km area is controlled by the net freshwater flux, i.e. by precipitation and evaporation. Here, precipitation is identical everwhere because it is not impacted by any ocean properties, while evaporation depends largely on atmospheric parameters like wind speed, relative humidity, and air pressure, but also on the air-sea temperature difference. The latter exhibits submesoscale spatial variability, but the impact on the surface salinity is negligible because the evaporated freshwater causes an increase of the surface salinity that is rapidly distributed within the mixed layer. It can be shown by a back-of-the-envelope calculation that under realistic conditions (evaporation rate ~ 1 mm/day, mixed-layer depth ~ 10 m), the corresponding salinity change is on the order of 10^{-3}. This may blur the submesoscale surface salinity distribution but the effect is merely detectable.

Action: We have added an explanation in Section 3 (new ms P6L31-P7L5).

(10)P7L5-6: No difference between the HBM and R500 on 1 June – it is the initial day when the fields are identical as seen in Fig. 2.

Action: Sentence changed (new ms P7L6-8)

(11)P7L12-13: How the domain averaged TKE is calculated (all model layers, volume average)?

TKE is read directly from the diagnostics in the ROMS logfile. Units are [Energy]/[Mass]=Nm/kg=kg m s^{-2} kg^{-1}=m s^{-2}

Action: A definition was added (new ms P7L14-15)

(12)P7L13-14: What is “the cumulative average TKER500”?

In a *cumulative moving average* or just *cumulative average*, the data arrive in an ordered datum stream, and the user would like to get the average of all of the data up until the current datum point (see https://en.wikipedia.org/wiki/Moving_average).
Example: Assume we have a time series of n values x1, x2, x3, …, xn.
The first element C(1) of the cumulative average is C(1)=x1
The second element is C(2)=(x1+x2)/2
The third element is C(3)=(x1+x2+x3)/3
…
The n-th element is C(n)=(x1+x2+x3+ … xn)/n

Action:

- A definition of the cumulative average was added (new ms P7L24-25)
- In order to avoid confusion, TKE was replaced by KE, because the acronym TKE is frequently used for *turbulent* kinetic energy.

(13)P7L15-25: The aim of the entire section is unclear. You wrote that the idea was to provide a rough estimate of the spin-up time, but you discuss something else. It is a numerical experiment to show what happens if you use the initial and boundary conditions with atmospheric forcing and finer model domain where the forcing is turned off.

Action: We have rewritten the entire paragraph (new ms P7L12-28)

(14)P7L26-32: From this, it is clear that the results presented later as R100 outcome are non-realistic (these are the results of an artificial numerical experiment; no point to compare the results directly with measurements).

Action: The entire paragraph has been rewritten (new ms P7L29-P8L8)

(15)P7L33-P8L8: This qualitative/visual comparison is OK, but I would not recommend to focus on it too much. A question would be whether R500 did reproduce the pattern qualitatively better than HBM, for instance.

Good advice!

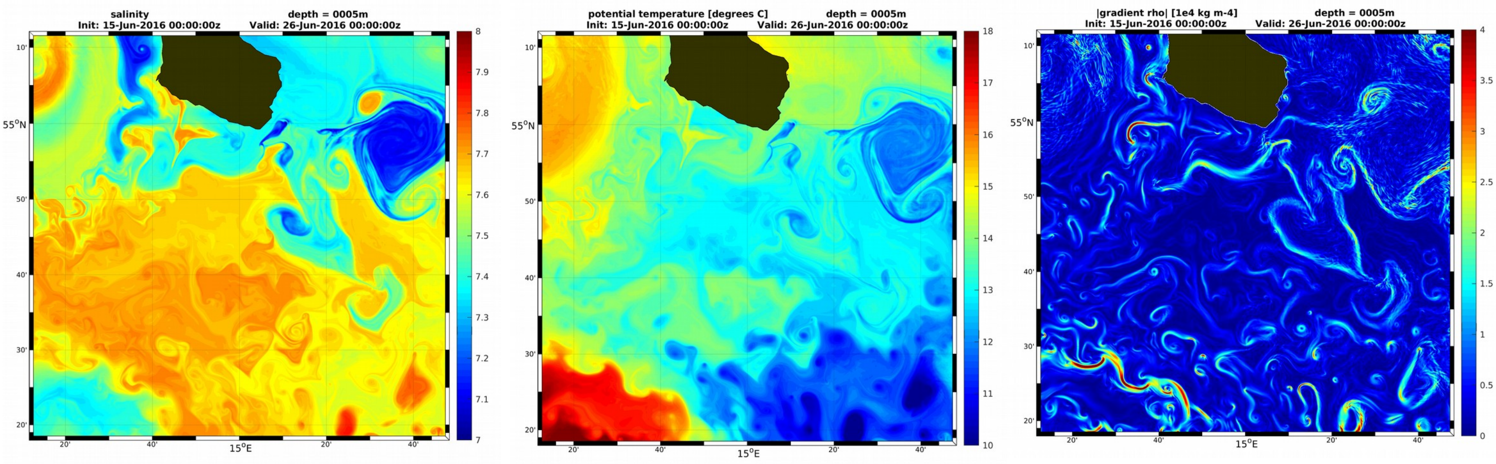
Action: HBM and R500 are compared in the new ms P8L19-21

(16)P8L10-15: It is a numerical experiment, where the forcing is turned off, and the aim could not be to reproduce the observed fields. A comparison with the measurements would be feasible only if forcing is on.

Action: To make it clearer, a piece of text was added (new ms P8L26-28))

(17)P8L22: A very thin surface layer is picked for the presented maps. Could it be possible to show the same maps, for instance, at 5 m depth?

Below are shown salinity, potential temperature and the absolute horizontal density gradient on 26 June at 5-m depth. Except for that the width of frontal zones is wider, these images resemble closely the corresponding images in Fig. 5.



<u>Action</u>: none

(18)P9L2-5: Is this effect caused by the fact that you have initial and boundary conditions taken from a model output with forcing and the maps presented are from a sub-region without forcing? It is not clear a priori what is related to the natural variability and what to the model set-up.

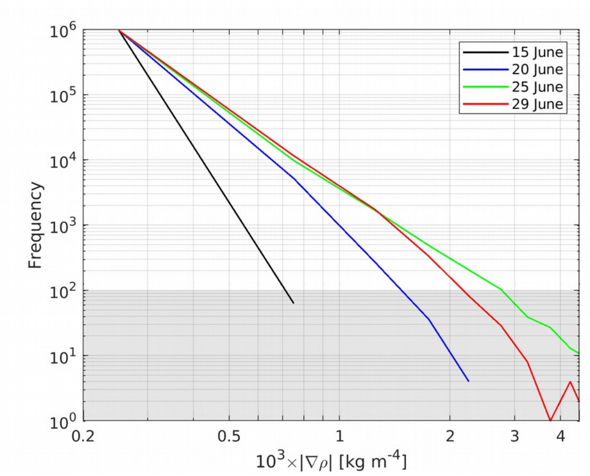
We have repeated R100, but using R500_NF for the initialization and the boundary conditions. The frequency distribution of $|\nabla\rho|$ is shown in Fig. B below, while the distribution of the run described in the manuscript is shown in Fig. A. One can see:

(1) On 15 June (=initial conditions), there are more high-gradient areas in B, or in other words, the gradients are stronger in B. This is plausible, because the atmospheric forcing in R500 has blurred the gradients.
(2) Same on 20 June
(3) Distributions on 25 June in A resembles closely the distribution on 20 June in B.
(4) 29 June: more high-gradient areas in A (in contrast to 15 June)
(5) In A, the frequency of strong gradients increases until 25 June. Thereafter, it decreases.
(6) In B, the frequency of strong gradients increases until 20 June. Thereafter, it decreases.

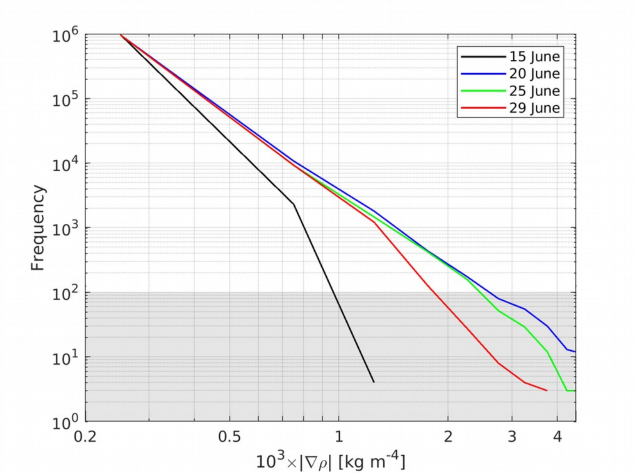
Hence, according to (5) and (6), the “frontal arrest“ occurs apparently when the strong gradients reach a critical value. Thereafter, the strong gradients become weaker. The critical value in A is reached on 25 June, while in B already on 20 June. It is not clear whether physical processes or numerical diffusion (or both) limit the increase of gradients. For the physical processes, Q_w, the straining deformation by the vertical velocity (see equation (7)), would be a suitable candidate. However, that would require to compute time series of the components of the tendency equation which is beyond the scope of this paper.

<u>Action</u>:

- These new aspects are discussed in the new ms P9L20-P10L2 and in P21L8-9
- We have added new Fig. 6



A: Frequency distribution of |∇ρ|. R100 was initialized from R500



B: Frequency distribution of |∇ρ|. R100 was initialized from R500_NF

(19)P9L13-14: In addition to dates, also the time should be referred (these figures are not daily average fields, I suppose).

Action:
A remark has been added that all figures are snapshots taken at 00:00 (new ms P9L3)

(20)P9L16-22: The same question as above – you should interpret the data as outcomes of a numerical experiment.

Action: This should be clarified by the piece of text added on P8L26-28 (see above **(16)**)

(21)P10L9-13: What that means? Could cubic spline cause such structures in the derivatives of the fields?

No – it‘s just the contrary! Without cubic splines, the structures look even worse. See above **(7)**.

Action: none

(22)P10L23-25 and P11L6-7: Could you reveal the period of these internal waves?

From a video of the vertical speed at 10-m depth in R100, we estimated the period of internal waves to τ=4 h=14 400 s which is equivalent to an angular frequency $\omega \approx 4 \times 10^{-4}$ s^{-1} . This is comparable to the buoyancy frequeny at the same depth of about 5×10^{-4} s^{-1}

Action:
As also the other reviewer asked for the properties of the internal waves, text was added in the new ms P11L15-21.

(23)P11L3: Which Class?

Class I – corrected.

Action: see new ms P11L33

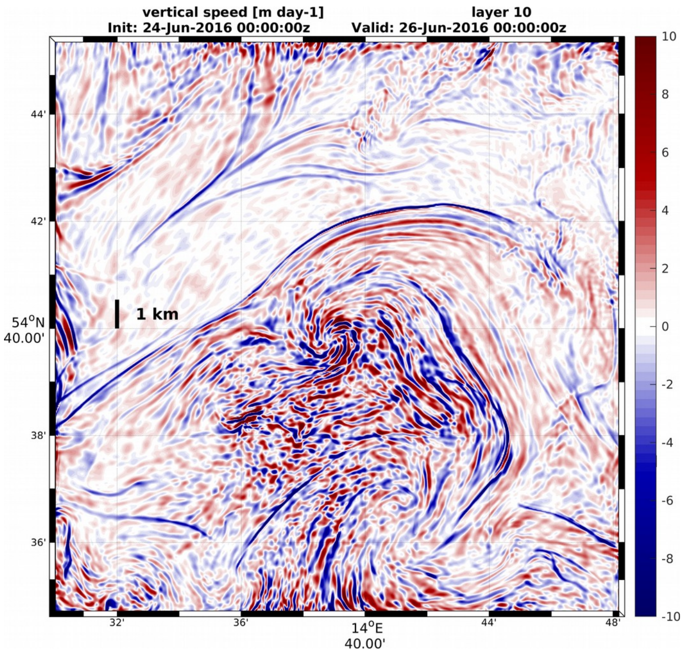
(24)P12L12: Fig. 9 top row is referred to as the vertical velocity at 5m depth, but in Fig. 9c the vertical velocity at the base of the top layer is presented.

Vertical velocity at the base of the top layer is correct.

Action: Text corrected (new ms P13L25)

(25)P12L29-30: Could the used cubic spline create more structures with the length scale of 1 km (R500 has the resolution of 500m)? Such a scale is well visible in many figures.

No, definitely not. The splines are used for the creation of the initial fields on 15 June in the entire R100 domain and thereafter only along the open boundaries. (Old) Fig. 9 shows the situation on 26 June, hence R100 should have “forgotten“ what has happened 11 days before. False advection from the boundaries is also unlikely, because the cyclone C3 is located in a quiet region. Meanwhile, we are investigating the life cycle of C3 in a third nest with 33.3 m resolution. In the figure below is shown the vertical velocity in that nest at the base of the near-surface layer. Also her, one can see wavelike structures with wavelengths of somewhat less than 1 km. These waves cannot be caused by the cubic splines because the grid size of the parent model, R100, is 100 m.



Action: none

(26)P12L31: Correct to “bimodal”. What are the three-modal structures?

Apparently, “three-modal“ does not exist in English.

Action: Sentence changed (new ms P14L16-17)

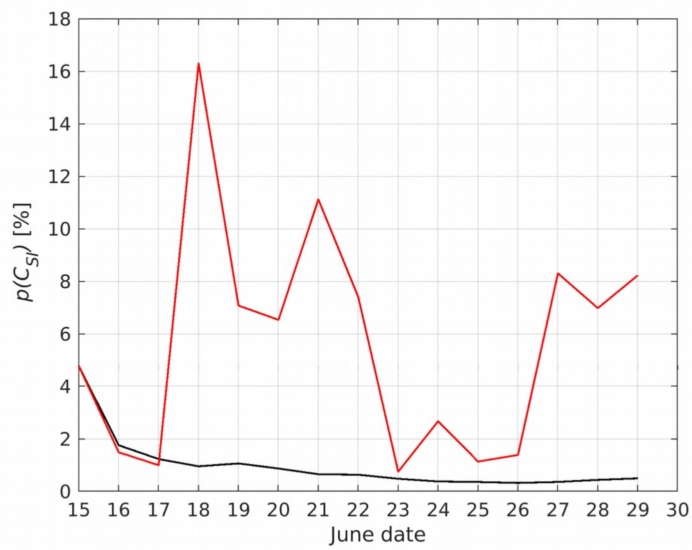
(27)P13L25: Could such false advection effects influence the results (statistics) in general as well?

In principle yes. However, as the region with false advection is limited and our results agree with theory (Haine and Marshall, 1998; Thomas et al, 2013), we do not believe that the error is significant.

Action: none

(28)P15L15-16 and P15L25-26: How these findings of fast changes during first days of integration can be interpreted? Are these caused by the fact that the initial fields were taken from a model output with atmospheric forcing and R100 was run without atmospheric forcing?

The rapid changes during the first days are in agreement with the theory of Haine and Marshall (1998): "This symmetric instability rapidly generates a layer with vanishing potential vorticity (Ri=1), but zero vertical stratification. … experiments indicate that the PV of the surface-forced layer is reset to zero on a timescale of a few hours". Nevertheless, we repeated the R100 run **with** atmospheric forcing, and the figure below shows p(C_{SI}) for the run without forcing (black) and the run with forcing (red). From 15 to 17 June, both curves are identical, but thereafter, p(C_{SI}) with forcing increases strongly and exhibits maxima on 18, 21, 24, and 27 June. These maxima are correlated with the wind bursts shown in Fig. 3. Hence, the conditions favouring symmetric instability are re-established during strong wind events by the buildup of potential energy and negative potential vorticity, and as soon as the wind slackens, the energy is released.



Action: We added a remark, see new ms P17L8-12.

(29)P17L26-28: Please, give references.

Action: The corresponding sentence was removed

(30)P17L29-30: It is true, if you mean 2D distributions. However, there are publications based on ferrybox and glider measurements covering large areas and presenting statistics of submesoscale variability (even in the Baltic).

If you mean the publication of

- Lips et al., Ocean Science, 12, 715 – 732, 2016: as we did not compute spectra, there are no quantities to compare.
- Karimova and Gade, International Journal of Remote Sensig, 37(10), 2394-2414, 2016: contains an eddy count of the entire Baltic; not suited for comparisons.

Action: none

(31)P17L32-P18L1: What do you mean by “hydrodynamical instability” here?

Fig. 15 and the corresponding pieces of text were removed

Action: N/A

(32)P18L9-10: Please, give references.

Fig. 15 and the corresponding pieces of text were removed

Action: N/A

(33)P18L14-15: Such 10m spots are not relevant in this context.

Fig. 16 and the corresponding pieces of text were removed

Action: N/A

(34)P18L21: Also Fig. 17 is not directly necessary to be presented.

Action: see below **(45)**

(35)P18L30-31: It is true in autumn-winter, but not in summer in the upper layer of the Baltic where the seasonal thermocline develops.

Action: The sentence was modified. See new ms P20L10-12.

(36)P20L8-10: This is a crucial point that the forcing was turned off in the high-resolution model domain, but it was still turned on in the model from where the initial and boundary conditions were extracted.

The purpose of the R500 run with forcing was to create a mesoscale environment that is as close as possible to reality for the entire month of June 2016. And the purpose of the R100 run was to study the evolution of STPPs in a realistic mesoscale environment. Therefore, the forcing was turned on in R500. As expected from previous studies - and also from our test runs – STPPs develop most rapidly when the (wind) forcing slackens, hence the forcing was turned off in R100 during its integration time 15 – 30 June.

Action: We tried to make it clear. See new ms P1L7, P4L8-12, P7L29-P8L8, P8L26-28, P9L20-P10L2, P12L14-33, P17L8-12, P21L5-7.

(37)P21L3-4: Is this statement about cyclonic eddies favoring the plankton growth a result of the present study? You could insert a reference.

Action: Reference to Mahadevan (2016) was added. See new ms P21L34.

Tables

(38) Table 1: I hope the number of vertical layers is more than 10.

The number of vertical levels is indeed 10 (ten). This is considered to be sufficient, because the water depth in R100 does not exceed 50 m. The vertical levels at the maximum depth are at about [0

1 3 5 9 15 21 26 32 39 50] m depth. Please note as well that the minimum water depth in the model is 5 m. There, the vertical levels are at [0 0.3 0.7 1.1 1.6 2.1 2.6 3.1 3.7 4.3 5.0] m. Further increase of the vertical resolution would require a much smaller time step in order to avoid violation of the vertical CFL.
By the way, in the 2016 version, the HIROMB model also has 10 vertical layers in the upper 50 m.

Action: none

Figures

(39) Fig. 2: What is meant by cumulatively averaged TKE? Where the wind data come from?

Cumulative average: see above **(12)**
Wind data: the wind data is the domain-averaged wind (from ICON-EU)

Action:
None, because in the caption of Fig. 3 is already written, that “All quantities are averaged over the model domain“

(40) Fig. 5: Could you add time (is it 0:00)?

Action:
The text “ All figures are snapshots taken at the specified day at 00:00“ was added in the new ms P9L3-4. See also above **(19)**

(41) Fig. 6: Time reference is missing.Fig. 8: Density anomaly values could be given.

Action:

- Fig. 6 “Time reference missing“: see above **(40)**
- Fig. 8: Density contours were labelled (see new ms Fig. 10)

(42) Fig. 11: Check the caption for English.

Action: done. See new ms Fig. 13

(43) Fig. 15: What is the location of this image? Is it relevant here?
Action: Fig. 15 was dropped

(44) Fig. 16: It is something else than discussed in the manuscript. Consider dropping this figure.
Action: Fig. 16 was dropped

(45) Fig. 17. I am not sure how relevant this figure is, especially since it is from the ocean while the paper is about the Baltic where the scales are different.

Action:
We would like to keep this figure, because it resembles closely the model eddy C3 and it exhibits details which may only be resolved by nonhydrostatic models.

We hope that our actions satisfy your criticism!

Best regards,
Reiner Onken and co-authors