We thank the Referee for his/her time and his/her constructive comments. We have complied with most of the proposed changes. In the following, the comments made by the Referee appear in black, while our replies are in blue.

Having gone through the revised manuscript and authors' response, I can see the revision is an overal improvement over the original and the scope or purpose of (some section of) the manuscript has become more clear to me now.

I would still like to comment on some of the responses and changes, as I generally feel the major concerns have only been minimally addressed. Apologies if I did not clearly convey my concerns in the original review.

**Regarding "Section 4 shortcomings"**
I appreciate the rephrasing of the text and the addition of Figure 10. Your goal of revised section 4 is more clear to me now, Figure 10 adds a form of validation compared to theoretical expectations and a comparison to a coarser reference, and it's now clear to me that in-depth quantitative validation and physical understanding is reserved for separate studies.

However, "physical realism" is not substantiated (beyond energy cascade at some level in the boundary layer) and still has to be believed by the reader based on the visualization of a few data fields (figures 9, 11), relative to AROME (figure 10), or relative to ECMWF-IFS (figure 11). Referring to separate future studies for validation is not sufficient. A "Devil's advocate" example: just because Figure 11b has a lot of detail does not mean the operational IFS forecast is worse, it's really a meaningless comparison unless there is a panel 11c showing (buoy?) data of wave height in various locations - for starters.

Ultimately, I'm simply not convinced about the purpose of section 4 in the context of the goal of the study. You have ported the model partially to the GPU and are now ready to numerically run at hectometric resolution. Section 4 successfully shows this, and if that is the goal, then OK. However, section 4 in its current form does not show that the model is ready to run at hectometric resolution in terms of physical realism.

We agree that Sect. 4 does not actually demonstrate "physical realism". The goal is in fact (i) to show that the code runs at hectometric resolution on very large grids and (ii) to illustrate possible weather applications. Validation with observations is not only beyond the scope of the article, but for the storms illustrated, it is likely more sensitive to the initial conditions (AROME vs. ECMWF for the example shown in Fig. 11) than to the detailed representation of physical processes.

Accordingly, we made the following changes in Sect. 4:

- Fig. 11 and its description were removed, while Sect. 4.1 and 4.2 were merged and slightly reorganized;
- "The physical realism and numerical efficiency of the Meso-NH code ported to GPU" was changed to "The numerical efficiency of the Meso-NH code ported to GPU";
- "[...] the results illustrate that different types of storms are realistically simulated by Meso-NH and benefit from the combination of large domain and high resolution" was changed to "[...] the results illustrate how the simulation of different types of storms may benefit from the combination of large domain and high resolution."

**Regarding "Reproducibility"**
I understand that you can't take each user's specific system into account. I was expecting more user-friendly instructions given the code's complexity, and it is hard to figure out what is going wrong if an error is in French or if a compilation issue does not give any information the cause of error. My concern was meant in part as honest user feedback (even beyond the scope of just this manuscript) and in part for being clear that I can't fulfil my role as reviewer in terms of checking reproducibility. I suppose I could indeed contact support, but I'm afraid we've then gone beyond what can be expected of me as a reviewer given available time.

We recognize that checking reproducibility is a hard task for a reviewer, if not impossible, since you do not have access to the supercomputers we used.
Regarding the French error code you mentioned in your first review, we have translated these comments written in French into English in the configure file. The updated version of the configure file can be found in a new zenodo repository.
Regarding the compile link issue you mentioned in your first review, it is very challenging for us to give answers without a direct exchange with the Meso-NH support (i.e. Juan Escobar and Philippe Wautelet).
To solve this issue, you will find in the new zenodo repository at `https://zenodo.org/doi/10.5281/zenodo.14872313` a Singularity image of a small test that we hope you will be able to run on one of the NVIDIA GPU workstations available to you.

**Regarding "Scaling with radiation"**
Radiation was just an example of where model resolution and physical parameterizations can go out of balance. I think my original comment was not phrased properly, but I believe it is important to address the physics of the model when going to such high resolution, and whether it makes sense to do so. Even for just the radiation-related questions, the only change I can see is that there is now a reference to a paper that details the radiation model, which I find inadequate:

- Given the description of other model components that are listed/referenced, you could have elaborated more on what kind of radiation you are running and what its limitations are.
- At what spatial resolution do you run radiation? How does it compare to hectometric resolution simulations? How do you generally justify running an older radiation transfer model for a global (then) 15+ km resolution scale model at a 100m resolution?

I think these are important themes in physical modelling and tie back to my concerns regarding "physical realism".

Regarding "physical realism" of our simulations, see above our response to your first comment.

The radiation scheme used in our study is indeed the old radiation transfer model of ECMWF, which is still interfaced with Meso-NH. As this scheme comes from ECMWF and this article is about porting the Meso-NH model to GPUs, no attempt has been made to port it to the GPU. In Sect 2.1, we stated that "the [radiation scheme] originates from the ECMWF and no attempt has been made to port it to the GPU". In Sect. 4.1, the weather applications we show focus on "the broad range of mechanisms involved in the formation of small-scale wind gusts in storms", in which radiation plays no or only a minor role. These two points explain why radiation is outside the scope of our article, and why we refrain from detailing any radiation-related issues you suggest.

We recognize, however, that radiation can be a key element in meteorological applications other than storms. This is particularly the case for fog and shallow clouds (see the two Meso-NH studies by Bergot et al., 2015 and Villefranque et al., 2022, cited in the introduction to the article). In the conclusion, we added "A physical parameterization not included in PHYEX is the ECMWF radiation scheme (neither the version by Gregory et al. 2000, nor ecRad by Hogan and Bozzo, 2018). The most recent scheme, ecRad, is a method for efficiently handling the 3D radiative effects associated with clouds, a property essential for fine resolution. It is currently being ported to GPUs at ECMWF and will be included, when available, in a future version of Meso-NH."

We thank the Referee for his time and his constructive comments. We have complied with most of the proposed changes. In the following, the comments made by the Referee appear in black, while our replies are in blue.

Thank you for addressing my comments. I still feel the need to highlight some minor issues with the replies and revised manuscript:

- In response to my former comment #1, in (3), you highlight a direct Poisson solver, not a conjugate residual algorithm, so the baseline numerical approach that is being used is not very clear to me. Is it an iterative method combined with FFT? Or is it a direct method that solves the Poisson equation in one step? Also, the decomposition you mention seems two-dimensional, not three-dimensional.
As written in the first paragraph of Section 2.1, "The current pressure solver consists on a conjugate-residual algorithm (Skamarock et al., 1997) accelerated by a flat fast Fourier transform (FFT) preconditioner (Bernardet, 1995)." To make the point clearer, the sentence is now "The current pressure solver is an iterative method consisting of a conjugate-residual algorithm (Skamarock et al., 1997) accelerated by a flat fast Fourier transform (FFT) preconditioner (Bernardet, 1995)."
In the third paragraph, we wrote "a 3-dimensional decomposition of the pencil was implemented and optimized." The decomposition is therefore three-dimensional.

- In response to my former comment #7, could you find a better reference explaining this approach? I had to download the repo and compile the latex file to see it.
Sorry, we could not find a better reference. To avoid downloading the repository and recompiling the latex file, we added GeometricMG.pdf, the compiled latex file to the new zenodo repository at `https://zenodo.org/doi/10.5281/zenodo.14872313`.

- In response to my former comment #10. I understand how there can be different binding configurations, but it is still unclear to me how more than one MPI task per GPU is implemented: Is there a domain decomposition within this device, or only several hosts managing one device? I would appreciate it if this could be made clear in the manuscript.
To be more explicit, the link between the domain decomposition on the MPI processes and on their counterpart on the GPUs was added with the following paragraph: "The domain decomposition is performed by dividing the domain into blocks of the closest possible dimensions in both horizontal directions and distributing the subdomains on the different MPI processes (one subdomain per process). Then, each MPI process is associated (bound) with a GPU on which it will offload part of its calculations (those that have been ported to GPU). It is then possible to associate several MPI processes with the same GPU which will then share its resources (in a way quite similar to sharing cores on a CPU)."