Dear Editor,

first of all I would like to say that I feel honored by the invitation to be a reviewer of the manuscript gmd-2022-204.

The manuscript entitled "How does 4DVar data assimilation affect the vertical representation of mesoscale eddies? A case study with OSSEs using ROMS v3.9" addresses a relevant and not entirely dominated question related to the quality of subsurface fields of ocean circulation models with observational data assimilation.

In general, the authors report their work in an organized, objective and well structured text, in a way that I could easily understand the problem, the methodological approach used, results obtained and what they indicate. The way the authors conduct the comparison of dynamic modes between solutions is particularly interesting.

I therefore recommend the publication of the paper after minor reviews are addressed.

Some questions, general comments, and suggested corrections are listed below so that the authors might want to address:

line 18) wouldn't be "to deliver" instead of "the deliver"?
although details of the DA model setup are reported in the other works cited, I missed some basic information of the 4D-Var implemented in ROMS that could somehow impact the results obtained, for example, the horizontal and vertical decorrelation scales, errors of the observations (table 1) and the number of inner and outer loops of the IS4D-Var. Has any sensitivity analysis been done in order to verify whether some of these assimilation parameters affect in a significant way the patterns found, i.e. the poor representation of the subsurface structures? Would reducing the errors of the observations bring the models closer to the reference state?

Do the fact that the reference run has a distinct setup with different boundary conditions and vertical mixing schemes, for example, interfere with the ability of the assimilative run to converge to the reference solution with respect to, for example, mixed layer depth? Could a distinct setup between the reference and assimilative model lead to biases that could not be corrected through DA? How do the reference simulation and the free integration of DA setup compare?

How was the density perturbation \( \rho' \) estimated? Was it calculated as the perturbation of a time and area average density \( \rho(z) \)?

Fontsize of vertical axis ticklabels and titles are too small

Colorbar labels are difficult to read on panels f) and I think a colorbar is missing for panels k) (green to red diverging colormap)

Is the word "sim" a typo in "sim350m"?

In this case, could the balance operator implemented in ROMS be favorable?