Dear Madam, Sir,

We thank 2 anonymous reviewers for the precise and helpful comments about our manuscript. We realize that a lot of time went into the reviewing process, giving the precise (and justified) remarks. We answered each of the comments, and modified the text and the figures accordingly. The revised version of the manuscript is much clearer, some mistakes was corrected, some new references were added, and some more links to CMEMS products were given, all according to the reviewer's comments.
In the replies to the reviewer #1, we announced to add a new annex with details of the data assimilation method. After completing also the correction of the manuscript for reviewer #2, we think that these details should be provided directly in the concerned section (instead of creating an annex). This is the only change in the reviewed paper, inconsistent with what we replied in the OSD forum. It is written in red in the text below.

We think the revised paper is much better than the original one, and submit it to you for consideration for publication in OS.

The replies to the reviewers were posted in the OSD discussion. They are copied below for your convience.

Reviewer 1
Specific comments:
1) My main concern is that "upscaling" appears to be feasible only if a similar setup is made between the child and the parent models. For instance, the authors use two models based on the same platform, i.e. NEMO, with an exact ration between horizontal grids and I suspect (not written in the text) with an identical vertical grid. All these are OK coinciding with the options to emulate two-way nested simulation. However, within a DA framework one would expect to see more general options, for instance, assimilating pseudo-observations on an entirely different grid (especially vertical for the T, S). The latter would support a more general argument for "upscaling" approaches, using for instance a different setup/grid/platform for the nested model. I leave it up to the authors choice if they wish to perform a DA experiment with a slightly different projection of the pseudo observations. However, I find useful the authors to discuss the limitations of their method.
The reviewer is entirely right that the configuration used to test the upscaling method is based on a nested grid setup using the same model code (Nemo) for both the parent and child grid; and furthermore the vertical grid is also identical (only the horizontal grid is different). This may influence the conclusion compared to a configuration with 2 different model codes. However we think that it is not a fundamental limit of the method a) Concerning the vertical grid, in the "normal" case of assimilating real observations, the latter are on a different grid than the model. Similarly if the child model was on a different vertical grid than the parent model, it would still contain useful information, and be worth to be assimilated in the parent model. What may happen however, is that some observations could be lost (e.g. the lowest model of the child model could be out-of-grid in the parent model)
b) if different model codes are used, the models could represent different processes. Hence, this should be taken into account by modifying the (representativity part of the) observation error covariance matrix. Examples of contributions to the representativity error could be
- different vertical coordinates
- different representations of the surface: rigid lid, free surface (with a linear or non-linear representation e.g. in Nemo)
- hydrostatic model, or not

- different atmospheric forcing fields
- different turbulent closure schemes
- different numerical schemes for advection, horizontal diffusion etc.

It is our opinion however, that between the parent and child models, the most striking difference is the horizontal resolution, and that therefor, the general conclusions of the paper are valid, and upscaling should not be limited to the case of parent and child models being identical. This is now better explained in the paper

2) page 2, line 16: "By upscaling the child model into the parent, the latter is brought closer to the former.". The benefits for the child model are obvious, though not so obvious for the parent model. Can the authors provide some guidance for "safe upscaling”? The way this work is constructed, suggests that a forecasting center should only "upscale" in case the child model has a similar modelling setup with the parent, e.g. same platform, vertical discretization, physics, parametrizations etc. The authors should also provide more information in the text about the setup of both models, in order to highlight their differences.
If one considers that the child model is better in its domain than the parent model (e.g. by comparison with real observations), then it would be desirable to upscale it into the parent model. This would be the case is some processes are dependent on resolution, in straits, etc; and is closely linked to the first specific comment in the review. We provide now a table in the annex of the paper giving details about the setup of both models; but upscaling should not be limited to identical parent and child models (see answer to comment 1)

3) page 4, line 23: "these pseudo-observations coming from the nested model are considered independent". This is a very strong assumption, since observations are on C2 OSD Interactive comment Printer-friendly version Discussion paper horizontal resolution 1/80◦ . In DA a common practice to avoid correlated errors is thinning or superobbing. Can the authors justify their option not to apply these techniques?
The reviewer is correct, that the assumption of spatially independant pseudo-observations is very strong. We are actually working on a non-diagonal observation error covariance matrix, but this is a large work that would not fit into the current paper. However, the assumption is partly alleviated by increasing the (diagonal) part of the matrix, in order to compensate for the (missing) non-diagonal elements. Increasing the diagonal elements in the matrix by a factor 3, for example, is similar to thinning observations with a factor 3. This is now stated in the paper.

4) page 4, lines 21-22: "Ensemble Kalman filter" and page 5, line 11: "Ensemble Tranform Kalman Filter variant of the EnKF". Use also in page 4 the word "variant". In addition, the authors should write in this section the DA method in more details. For instance, it should be mentioned that this is a deterministic approach of the EnKF, i.e. pseudo-observations from the child model are not perturbed and the perturbation approach is only applied in the parent to obtain model errors. All these are not apparent to the reader, at least not before start reading the results section.
We added an annex to the paper with the details of the data assimilation filter used in the study.
In the final revised version of the paper, we decided not to enter this information in an annex, but directly in the paper, as suggested by the reviewer.

5) page 6, lines 13-14: "to update directly the tiles from the Mediterranean model restart files, influenced by the nested model, without including the other tiles in the state vector". This is an interesting technical capability of OAK, but if not mistaken that means that there is a crude correction cutoff in the neighboring tiles just outside the nested domain. I would assume that the localization is

enough to constrain the correction in an area slightly broader than the nested domain. Can the authors clarify what is the purpose of this capability?
In the state vector of the parallelized parent model, we include the tiles covered by the nested model, but also the tiles immediately around that area. Therefore, as the reviewer correctly assumes, the correction is not cut off at the margins of the area covered by the nested model, but propagates outside. The extent of the correction outside the area, depends on the radius used in the localization method. This is now better explained in the paper.

6) page 9, lines 7-8: "The ability to ... would be beneficial to constraint the model". This is more a concluding remark, rather than a result of the study. The phrase should be moved in the Conclusions section 5.
We moved the remark to the conclusions

7) Figure 5. The SST is L4 or L3? In section 3.4 it is mentioned as L3.
In the study (section 3.4), L3 satellite images are used. Only in figure 5 is the L4 image used for visual comparison of model and satellite image. This is now clarified in the article.

8) Figure 8. The units are missing from the axes.
added units

9) whole page 16: "Advantages of using upscaling include ...". This is a nice summary of "upscaling" advantages supporting the method. Can the authors provide possible disadvantages (if there are any) and suggest possible remedies?
The reviewer is right that the list of advantages should be accompagnied with a list of possible limitations (or disadvantages). This is now included in the article, and copied here:
a) the child model should be "better" than the parent model
b) exactly as when assimilating real observations, the data assimilation procedure itself uses approximations, and this could degrade the analysis
c) if the parent and child models are very different, the parent model could not manage to ingest the pseudo-observations
d) the coupling with upscaling is not as strong as with real two-way nesting

Potential remedies for limitations b and c would be
b) see all the research about this problem (in the context of assimilation of real observations), such as anamorphosis techniques (when a non-linear relation exists between model variables and observations), particle filters (when the error distribution cannot be considered Gaussian), etc
c) the observation error covariance matrix should be specified carefully to filter out the processes of the child model, that cannot be represented in the parent model

Reviewer 2

Abstract

Line 5: *Therfore* instead of therefor and I would take out "*in practice*"
Changed to Therefore

Line 6: …" to replace the missing model feedback…" I would insert "*child model or high resolution model*". Done

Line 10: I suggest to rephrase something like:
"A basin scale model simulation is compared to one simulation…, and another model analysis which applies the upscaling technique…"
Changed

Introduction

Line 15: "reanalyses, analyses and forecasts" Changed
Line 16: " …by different institutes within the regional monitoring and forecasting centers.." Changed
Line 24: could you insert a reference for this? We rephrased the sentence instead
Line 6 page 2: I would take out "in this artice" Removed

Line 12 page 2: I would substitute *basin-scale* with *regional* Changed
Line 13 page 2: "…in the basin-scale model, … is to obtain" Changed
Line 14 page 2: "(along with…)" Changed
Please rephrase the entire sentence, it seems too informal to me Rephrased
Line 16 page 2: I suggest *"…to the child model will progressively gain consistency with the child model solution within its domain, being beneficial for the child model over time*." Changed
Line 20 page 2: do you have references for this? Added Mason et al 2010, Debreu et al 2012
Line 21 page 2: Is this pertinent? I do not see the connection, please explain. If we assimilate pseudo-observations coming from a nested model, and this improves our (parent) model, then the child model is a pseudo-measurement device which can be seen as a replacement for (costly) real measurement devices.
Line 26 page 2: Other re-initialization techniques have been used blending, through optimal interpolation, coarse resolution operational analyses and coastal observations in so called Rapid Environmental Assessment experiments. Please give a look at *Simoncelli et al.(2011)*, they show improvements in the nested coastal model performance using observations. Added
Line 35 page 2: I am not sure that the syntax is correct please check the English. Changed
Line 3 page 3: "Therefore acknowledging that operational…, Schulz-Stellenfleth and Stanev (2016) strongly…" Changed

2.1 Hydrodynamic Model

Line 13: I would substitute *tried out* with *has been implemented* Changed
Line 15: *"created by the junction of the Eastern and Western Corsican Currents"* I would cite *Pinardi et al (2015)* Added
Line 17: I would add some recent references *Pinardi et al (2015), Somot et al. (2016), Simoncelli and Pinardi in von Schuckman et al. (2018).* Added
Line 21: The resolution of MFS is 1/24th (Clementi et al. 2017, https://doi.org/10.25423/cmcc/medsea_analysis_forecast_phy_006_013) of a degree since October 2017. The reanalysis (https://doi.org/10.25423/medsea_reanalysis_phys_006_004 Simoncelli et al., 2014, 2016) is still at 1/16th. Changed

Line 30: please specify for reproducibility issues which analyses has been used, I guess the 1/16? Or the reanalyses? Please clarify and insert the reference. Added

Line 32: ERA Interim is not at a resolution of 1/8th of degree, it is 0.75 degrees!!! You might have re interpolated it from 1/8th to 1/16th. Please explain it and add the Dee et al () reference of ERA Interim.
Modified and added reference

Line 1 page 4: Which literature? Removed « from the litterature »
Line 2: 5 rivers' data comes from? Various regional websites, not added in the article
Line 4: Please consider the CMEMS has 39 rivers, thus it is not very much coherent. Please look at the CMEMS products descriptions and may be cite the reanalysis instead. Rivers are described in the Simoncelli et al 2016. Added in the text that CMEMS has many more rivers

Figure 1: Please increase the axis font, not readable now. I would show the two models' salinity fields to emphasize the differences due to the daily river outflow, instead of the difference. In fact, you describe the different plumes in the manuscript.
We tried this out, but due tu large salinity variability, a side-by-side plot of salinity does not show clearly the salinity difference. Therefore we chose to show directly the difference

Line 10: please describe more in detail Fig1b eventually, isn't it the difference among the two models' salinity after 1 month of simulation? Please improve also the caption.
Yes Fig 1b is the difference of (nested) model salinity after 1 month, when using climatological Rhone discharge or real, daily discharge. This is now written more clearly in the caption

2.2
Line 16: I would re-phrase something like "*In order to assimilate* …, different set ups could be implemented (adopted, applied) depending on …" Changed

Line 21: please improve the description of the settings, it contains repetitions.
Line 23-25: Start a new phrase please and please say something more about the statement that the T and S pseudo-obs are considered independent. You mean that you assume that even if it is not the case. You assimilate the full resolution 3D T and S fields? None thinning? Please motivate a bit this part. The word *also* could be neglected and substitute "*;*". It looks like you wrote this in a rush without much care. Improved lines 21-25 as suggested
As the reviewer correctly supposed, we do perform thinning (in the horizontal), and this is now also written in the text. We apologize for forgetting to write it in the article first submitted

Line 31: random not randon Changed

Line 7 Page 5: Why did you select 1 month of spin up time? Please start a new phrase and integrate a bit on that. What you use in the evaluation is the ensemble mean of the 100 members?
In a previous (single) model spin-up, we noticed that the kinetic energy reaches a more-or-less stable value in just a few days, therefor a spin-up of 1 month was considered sufficient. The same is supposed for the ensemble members. As we are now running an ensemble of 100 members, longer spin-up would translate into large computational cost.

Line 1 Page 6: it is not clear to me "*...and its addition observation localization*", could you please explain?

It is explained just afterwards (lines 2-3. In means that the model domain is cut into subdomains (water columns in our case) where the analysis is performed separately. Moreover, in each subdomain, only relevant observations are considered (i.e. observations that are far away and will have no impact, are not considered during the computation).

Line 7: many are the sensitivity experiments with different observation errors, maybe you could insert a table.
We agree that there are 4 experiments (each with a value for temperature and salinity observation errors), but these all fit into 1 line in the text. Maybe it is easier to keep them in the text ?

3. Metrics

Line 17: Please add something about: .. these metrics have been computed to compare model solutions. What do you want to show? Upscaling model solution (is it an ensemble mean of the 100 members?) with the basin scale simulation?
Added some explation to Line 17. About the « upscaled » model, we keep one unperturbed ensemble member. The other 99 members of the ensemble are perturbed and are used only to create the model error space. All 100 members are updated daily by the data assimilation procedure.

3.1
Line 23: see previous comment. What do you want to show? Does this metric tell you if the upscaling procedure is driving the upscaled model solution towards the NW-Med? Please integrate a bit.
Yes exactly, we want to see if the upscaling procedure pushes the parent model towards the NW-Med model. This is now explained better at line 23.

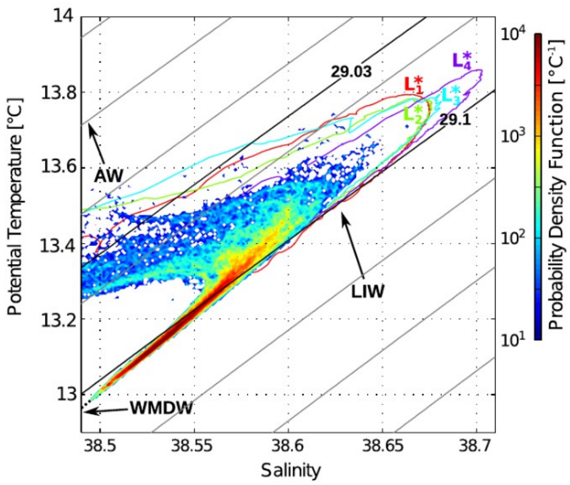
3.2
Line 28: again, see previous comments. I would integrate at the beginning of the section, and in each sub-section I would state what exactly is expected from each metrics.

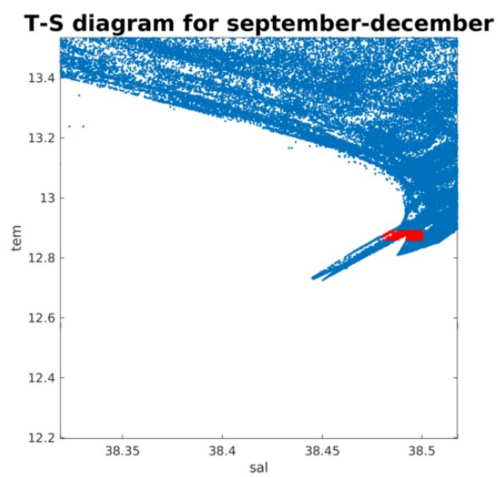
3.3
I do not really understand what you are going to present in the results. Plume length/direction from the free and upscaled models? The comparison model-obs using sat chlla images? “Furthermore,…” What does it mean you do it qualitatively, quantitatively, both?
Yes we compare plume length and directions in between models, and also between model and real observations. This is different compared to the previous metrics where we did not have real measurements. This is now better explained in the text.

3.4
ok

3.5
Line 15: What do you mean bay “depth is larger than 1000m”? Is it the depth reached by deep convection? Maybe you want to put it at the end of the first phrase at line 13. Here you are describing the WMDW characteristics.
Following Bosse at al (2015), we consider that water is WMDW if it is within a certain T and S range, AND if it is situated below 1000m depth

Line 17: Which tail? How could the river see that? Please explain or describe with some detail.\
Bosse et al (2015) show the T-S diagram and indicated the water masses (left plot below). We reproduce this plot from model results (right plot below) and obtain a longer tail (it goes to ~12.7°C and slightly lower than 38.45 psu).The definition from Bosse at al (2015) would result only in the red rectangle in the right plot below. Therefore we slightly adapt the WMDW definition in order to capture the whole tail. We decided not to include these 2 figures in order to keep the paper more consise.





4. Results

Figure 3: please enlarge the font. Changed
Line 5: The difference of what? Temperature, added in the text and in figure caption
Line 9: Isn't it Fig.1b the difference between unperturbed parent and child model in salinity after 1 month of spin up? Fig 1b is the salinity difference of the nested model, when forced with climatological or with real Rhone river discharge data. Figure 3 is the difference between parent and nested model (for temperature), and the equivalent for salinity is not shown in the article
Line 10: Why should I trust the child model more than the parent? You did not provide any model performance. It is not easy to improve significantly the smooth solution of the coarse parent model (phase errors are common).
Also following recommendation from reviewer 1, we now stated clearly that we *suppose* from the start that the nested model is better (in some sense) than the parent. We do not try to prove the hypothesis is actually valid, and sometimes, it could be wrong. Exactly as the reviewer points out, when the nested model represents small scales that are actually out of phase with reality, it could have higher RMS errors than the parent model that does not represent small scales at all.

But supposing that the nested model is « better », our objective with upscaling is to bring the parent model closer to the nested model (i.e. emulate nesting feedback). This is now written in the introduction.

Figure 4: please include a line to indicate the section location on the map. Done

Line 16: The scope of Figure 5 is to show that the upscaling technique is bringing the upscaled model close to the NW-Med one. RMS difference could be provided towards the satellite SST to show this. However, this rise the question: “ why don’t you consider to weight the assimilation of pseudo observations according to the misfit with the observed SST, giving more weight where the pseudo-obs are loser to sat obs? Please keep the same notation to call the models (MED, NW-Med…).
This is an interesting suggestion but has also the following limitation. If we have many observations to validate the nested model, then we could as well assimilate these (real) observations into the parent model. Upscaling is interesting mostly when there are few or no real observations.
In the case of SST however, it is a very interesting suggestion. When the nested model is well-validated by SST, maybe we can trust it more also for other variables and assimilate those in the parent model to complement the (real) SST observations. This could be the topic of a follow-up study.

Please switch 4.1 and 4.2 (thus fig. 6 and 7) to be consistent with 3.1 and 3.2. Done, we rather switched 3.1 and 3.2

4.1

Line 21: For which models? Parent model, free and upscaled. Added in the text

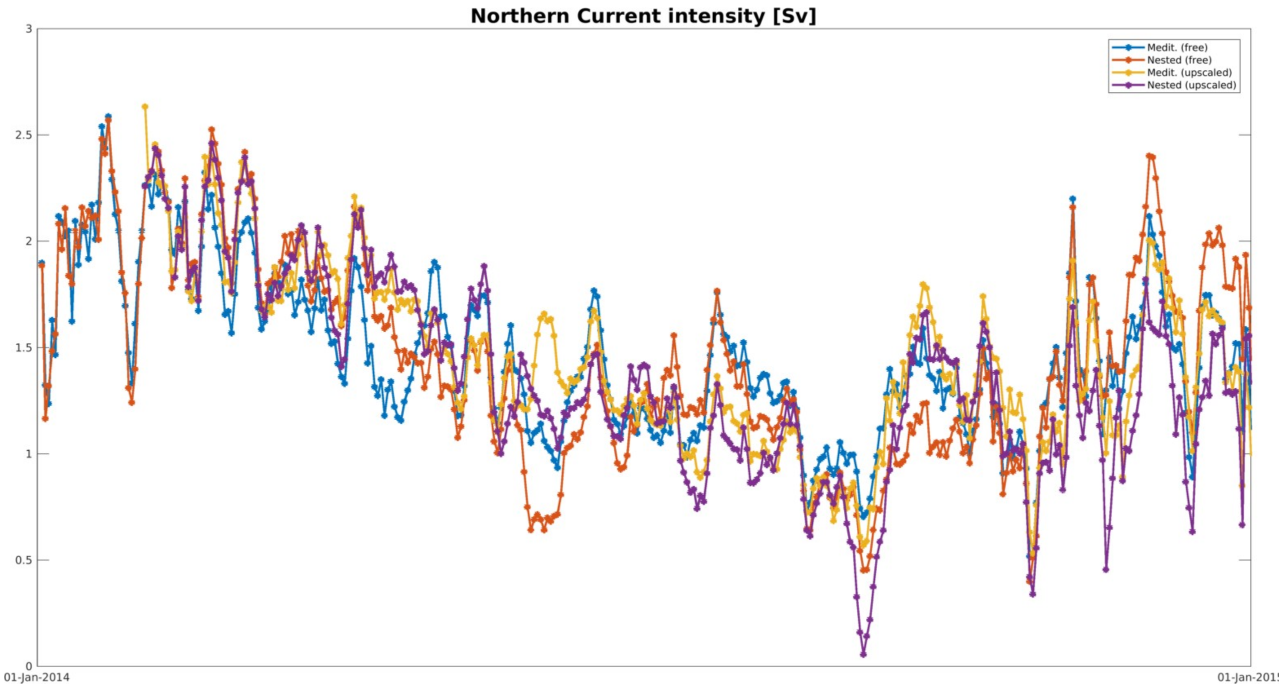
Line 2 Page 9: Why the NW-Med model is not shown in Figure 6? Please include it for consistency in presenting the results, since you want to prove that the upscaled model is driven towards the NW-Med solution. Again please keep the same nomenclature to facilitate the reader. Moreover RMSD could be computed as in 4.2 case.
Please see the answer to comment 4.2 (Fig. 7)

Lines 9-12: What is this meaning? Is this correct? Is this coherent with NW-Med?
Yes it can be seen from the figure that both lines are closer one-to-the-other after the first month. There is also a temporal delay in between the 2 curves in August-September. The text merely describes this.

Figure 6: please increase the font in the legend and include NW-Med.
We will modify the figure fonts

4.2

Figure 7: Please insert NW-Med and the legend.
Once upscaling modifies the parent model, the child model gets different open-sea boundary conditions from the parent, and starts to be different as well. Thus, for Figures 6 and 7, we could add 2 NW-Med curves to the figures (one corresponding to the free run and one corresponding to the upscaling run). For the Northern Current intensity, it gives this :



This figure becomes difficult to interpret. There is no general trend, like for example a systematically more intense Northern Current in nested models. However, what's interesting is that generally speaking, compared to the free Med model (blue curve), the yellow curve seems to get closer to the purple (child model of the upscaled Med model). The discrepancy between yellow and purple is smaller than the discrepancy between blue and red. This can be verified by looking at RMSD errors between parent and child model in the two cases (free model or upscaled model), respectively 0.22 and 0.19 Sv. The RMSD for all metrics have now been included in a table as suggested by the reviewer.

Lines 4 -6Page 11: I suggest to rephrase *"…this metrics cannot **be** used to compare and validate the models since observations are not available to compute the real NC transport. However, it shows that upscaling of.."*

But the metric *can* be used to inter-compare models, so it might be confusing to say « this metric cannot be used to compare … ». Therefore we rephrased as :

« For the purpose of our study, this metric cannot be used to validate the model since real measurements of the Northern Current transport are not available; but (as for the previous metric), it can be used to compare models, … »

4.3

As for the other metrics I would include the NW-Med to show its consistency with the upscaled model. At Line 15 you mention the nested (NW-Med) model but it is not shown. RMSD could be computed again as more robust argument of your results.

On Fig. 8, the arrows for the upscaled and (not-shown) NW-MED models superpose and are indistinguishable

Line 17: a quantitative estimate of model performance can be provided again by RMSD of MED, and upscaled model towards NW-Med.

We have computed the RMSD and included it in a table

Figure 9: This figure presents MED and NW-Med, why not the upscaled model? I suggest to shaw the three salinity fields.

It is indistinguishable from the nested model

Line 3 Page 12: Is the increase in salinity observed, is it consistent with observations or only with NW-Med? I would include validation with observations, since a lot of them should be available by REP in situ observations fromCMEMS.

We agree that it could be checked whether it is realistic or not. However even if it is *not* realistic, if the nested model predicts it, and upscaling can make the parend model predict it as well, then upscaling is doing what we hoped. Our hypothesis is always that the nested model is « better ».

In other words, if nesting predicts a saltier Corsican Current core, but it is actually not true, what do we hope from upscaling ? That the parent model also increases salinity in the core, or not ?

For this reason, we actually did not check if the nested model is more realistic or not. There are lots of other papers showing the impact of nesting.

4.4

Figure 10 should be mentioned since the beginning of the paragraph to help the reader.

Done

Please harmonize the models' nomenclature in the text but also in the legend and caption of figure 10. The validation (actually the only one that is provided with observations) should cover the entire 2014 year and not only two months.

In the text is written : « A similar plot for the whole of 2014 shows that the situation worsens during summer (errors of 3 ◦ C) both for parent and child model; the difference in between models is hidden by the temporal variability of the error (not shown). »

If we plot the entire year, due to the scale going to 3°C instead of 1.4°C, we would almost not distinguish the red, black and blue curves anymore at all.

The mean RMSE of the three models should be provided as well to support your results. Upscaling should consider the performance of the child model and assimilate only, or give more weight, in those parts of the domain where the child model is close to observations. Without considering models' performance upscaling could force the parent model towards a wrong solution.

This comment is related to a suggestion higher in the review. We agree that this could be a way forward for SST, but maybe in a follow-up study. Apart from SST, for the other metrics, there are no (real) data to choose whether or not to trust the nested model. The hypothesis of this study is that the nested model **is** better, and the objective is to bring the parent model closer to the nested model.

Lines 6-8 Page 13: why not? REP in situ temperature and salinity profiles are available from CMEMS. I encourage to look at them and compute some validation to make your paper more robust.

This comment is again related to the previous one, and other ones before. In this study, we **only** aim at bringing the parent model closer to the nested model. Validating the nested model, such as doing a QUID for it, is out of our scope (and a whole lot of work).

On this occasion of in-depth temperature, we remembered the reader of the hypothesis and objective of the paper.

Line 8: "Upscaling is able to bring the differences back to the parent model" or the child model?

We rephrased this so that is it more clear

4.5

Here you mention also the child-model of the upscaled model and you speculate without any proof about the largest consistency in the nested upscaled models. I recommend to include a figure to show it. Or if you want to use Fig. 11 please give the RMSD of the model couples or include a specific comment to the figure (Line 7-8?)
We have added a table with the RMSD of model couples, as suggested by the reviewer

Line 10: Therefore Changed

5. Conclusions

Lines 4-9 Page 15: “The underlining hypothesis…” You only state it in the conclusions while it should be written clearly both in the abstract and the introduction. In this way the reader is aware that you only aim to relax the child model solution to the parent model solution, independently from the model performances.
Yes exactly. Sorry for this. It is now written in the introduction

Line 13: please use child instead of nested. changed

Line 15: Please be more precise and refer to the figures when this is shown. changed

Lines 1-6 Page 16: This general statement might confuse the reader since your child models never assimilate observations. Please rephrase to underline this. Rephrased

Line 6: please include a reference since you do not show this in the paper. Changed the text

Line 7-end: The advantage might be real is the upscaled solution has a better skill towards observations. In the last phrase: If the high resolution model is upscaled into the basin-scale one? Please notice that in CMEMS regional systems are the basin-scale ones.
We added explicitely in the text the potential disadvantages of the method, which include the fact that the nested model may be worse than the parent.