We thank the referees who have made excellent work in going through the details of our submitted MS and made very constructive remarks and corrections. Our detailed step-by-step responses to each of the Referee #3 comments or questions are given below.

We have revised the MS, with the following main points.

- The main points of the EOF reconstruction and the found modes were presented too briefly, relying mainly on the reference Elken et al. (2019). In the revised MS, additional important issues have been included in the compact form (hopefully not repeating the already published MS).
- Justification for the large-scale EOF DA method, in comparison with other well-known DA methods, has been refined.
- Data transformations between the fine and coarse grids have been more carefully presented.
- Unfortunately, the issue of observational errors has not been included in the initial MS. It is now included in the revised MS.
- Presentation of DA validation has been reformulated and discussed in more details.
- Possibilities of the method regarding operational forecast (with assimilating only the past data) have been discussed.

Suggested technical corrections have been included as well.

Anonymous Referee #3

Received and published: 16 July 2020

Comments and questions in bold
Response by the authors in normal
Line and Figure numbers taken from first submission

General Comments
The paper addresses an important issue, which is the estimation of salinity and temperature for the Baltic Sea using a combination of model data and observations.

The method is based on a two-step approach, in which sparse observations are interpolated using an EOF technique and subsequently a relaxation method is applied for the assimilation into the model.

The method seems to have some potential for the assimilation of FerryBox data, where interpolation to 2D grids make sense, if longer time scales are considered.

There are a couple of concerns, which should be addressed:

- **The method assumes in the interpolation step, that the covariance structure of the model is correct. This should be discussed more – in particular the limitations caused by this**

Problems of EOF reconstruction have been considered by Elken et al. (2019). We cite: SST and SSS results are rather well validated by observations and the model-based covariance patterns can be considered trustful. /// Fu et al. (2011) compared covariance patterns from modeled SST and satellite SST, and found them agreeing well. CMEMS QUID report has presented validation of SSS against FerryBox data, showing that the SSS patterns were well simulated by the model. In deeper layers, however, there is usually a larger spread between different model results.

Main differences between actual and model-based covariance estimates are expected within very short term variations (occurring above the Nyquist frequency/wavenumber) that comprise in observational datasets spatially uncorrelated noise, using the terminology of optimal interpolation.

We have added following paragraph after line 487:
"We have tested the EOF-based DA in centred time window of 30 days, based mainly on available FerryBox data during the study period. As shown by reconstruction experiments by Elken et al. (2019), the time-dependent method can also work with backward observations as if it occurs during operational forecasts. When more observations become available, for example from new automated buoy stations, Argo floats and gliders, the time window can be shortened. Full covariance matrix estimated from the model results is the backbone of the EOF DA method. Prior and/or complementary to implementation of the method into operational practice, detailed covariance studies using results from multiple models could be useful, as well as additional reconstruction and DA studies using more data sources over longer periods."

- **The previous point is related to a discussion of the main model error sources, which is missing as well**

The model results, accuracy and error problems have been considered by the larger CMEMS community. Unfortunately, references were missing in the model description part, although they were in other places (Golbeck et al., 2015; Hernandez et al., 2015; Tuomi et al., 2018; Huess, 2020; She et al., 2020). They have been added in the section 2.1 of the revised MS. Text on lines 92 is extended to:

"Detailed description of the HBM model and its validation can be found by Berg and Poulsen (2012); <u>further analysis and evaluations are given by Golbeck et al., 2015; Hernandez et al., 2015; Tuomi et al., 2018; Huess, 2020; She et al., 2020. In particular, the CMEMS Quality Information Document (Golbeck et al., 2018) concludes that temperature forecast between the surface and about 100 m depth is one of the major strengths of the CMEMS-V4 product, below the halocline deviations of forecast from observations increase. Regarding salinity, the values are slightly underestimated and the underestimation increases with depth."</u>

- **Observation errors are not discussed at all – this needs to be justified and discussed**

Indeed, this important question was missing in our presentation. In meteorological terminology, our method is "analysis nudging" (e.g. Stauffer and Seaman, 1990) that makes Newtonian relaxation to the gridded fields reconstructed from the observations. The issues of observation errors are included in the reconstruction procedure, when values over (usually very small) sensor space are converted to the values over larger grid cells. DA based on the analysis nudging treat observational errors usually by adding appropriate white noise to the input data, before producing the gridded field to be used in relaxation. In this context, we think we have to make additional study on EOF reconstruction of noisy observations, in order to extend the first results presented by Elken et al. (2019). In this MS, which main focus is on computationally extensive model runs, we add several notes on the problem of observation errors.

Text on lines 164-166 is extended to:

"When L most energetic modes are taken into account in the sorted list of eigenvalues and -vectors, the sum from λ_1 to λ_L presents the explained variance and contribution of truncated modes forms the error variance. If white noise with a variance ε^2 is present in the decomposed data due to sub-grid scale processes and/or sampling errors, the noise variance appears only as additive to the diagonal elements of the covariance matrix. The eigenvalue problem becomes $(\mathbf{B} + \varepsilon^2\mathbf{I})\mathbf{E} = \mathbf{\Lambda}\mathbf{E}$, where $\mathbf{I}$ is a unity matrix. Patterns of spatial modes remain unaffected by adding the white noise, but the eigenvalues and energy share of the modes decrease according to a factor $(1 + \varepsilon^2/\sigma^2)^{-1}$. When the sum of eigenvalues of the included dominating modes is less than $\sigma^2 - \varepsilon^2$, contribution of noise is effectively smoothed."

Text on lines 217-218 is extended to:

This is the main DA calibration parameter, since extensive use of covariance statistics, including the effects of observation errors, has been included in the estimation of gridded reconstruction of point observations. Newtonian relaxation of gridded observations, applied during the model run at DA time steps is named also "analysis nudging" (e.g. Stauffer and Seaman, 1990), which has recent meteorological applications (Bullock et al., 2018).

Section 3.1.1 has been added:

START

3.1.1 Covariance, modes and reconstruction tests

The EOF modes were calculated on the coarse grid (5' N × 10' E) on the basis of space-averaged results from the fine grid (0.5' N × 1' E) model, running from 1 July 2010 to 30 June 30 2015 (Elken et al., 2019). This analysis revealed that mean distributions of modelled SST and SSS, serving as the basis for calculation of deviations in the variability studies, were close to the climatological maps calculated on the basis of observations (Janssen et al., 1999). Highest temporal variability was found in the shallow coastal areas for SST, whereas largest SSS variations were revealed near the larger river mouths and in the NE area of the Gulf of Finland. While temporal changes strongly dominate in the variability of SST, spatial changes prevail in SSS variability.

Calculated SST and SSS covariance matrices have significant spreading of individual values over pairs of points, especially for the dominating gravest modes where big covariance values may occur over large distances. Covariance of residual fields (sum of higher EOF modes) has a decay scale about 30 km with increasing space lag, both for SST and SSS. The first, most energetic EOF modes have nearly "flat" patterns without sign change (energy share 97.6% for SST and 36.2% for SSS); their amplitudes are dominated by a seasonal signal. Space-dependent mean biharmonic seasonal cycle was not removed from the model time series prior to the analysis, since special experiments revealed only a small effect of seasonality suppression on EOF mode patterns. Second EOF mode of SST (1.3%) presents differential heating and cooling in shallow areas, compared to the deeper offshore waters. Transverse anomaly stripes near northern or southern coasts, like due to coherent upwelling and downwelling in the region, were evident in the second SSS mode pattern (16.9%) and third SST mode pattern (0.31%). There is also a pattern of SSS changes in the freshwater spreading pathway near the northern coast of the Gulf of Finland (third SSS mode, 7.1%) and longitudinal SST changes in east-west direction (fourth SST mode, 0.14%).

The data set used in the present DA study (Fig. 2) is rather irregular, compared to the reconstruction experiments by Elken et al. (2019). Therefore, we revisit the covariance issues and perform additional reconstruction tests, before finding in the next subsection the best options for the automatic reconstruction procedure. Spatial interrelation of observed values at a specific point to the values in the rest of the region is found from the extract of the spatial covariance matrix, which can be shown as a map. One example of SSS covariance with a frequently sampled HELCOM monitoring station BMP F3 is shown in Fig. 3. The covariance of three dominating EOF modes (Fig. 3b) comprises most of the unfiltered data covariance (Fig. 3a) at large distances. High covariance locations have clear basin-scale geographical explanations: under the similar weather and seasonal forcing, which is spatially nearly uniform, SSS changes in distant river influence areas are closely interlinked. Correlation (not shown) may exceed 0.4 at distances greater than 500 km; therefore, assumptions of fast decay of correlation with space lag (like using the Gaussian covariance approximation), adopted in offshore areas with negligible coastal influence, are not valid. Covariance of residuals to the large-scale variations are presented by higher EOF modes (Fig. 3c). Such smaller scale variations have nearly Gaussian structure, with elliptical anisotropy stretched along the axis of the basins similar to the results by Høyer and She (2007): spatial scales in Fig. 3c are 30 km and 15 km along the main axis and perpendicular to the axis, respectively. Similar regularities – physically explained high covariance at large distances, localized covariance patterns for the higher EOF modes – were found for other points of reference, both for SSS and SST fields.

EOF reconstruction method relies on the full covariance matrix, without any approximation. Full covariance matrix can be implemented in optimal interpolation as well. While EOF method needs to limit the number of included modes, smoothing in such way smaller scale variability and observational errors, optimal interpolation needs to include observational error variance ("nugget effect" in terms of Kriging method, equivalent to optimal interpolation); otherwise the system of underlying linear equations may become close to singular and the result may become unrealistically spiky. In some examples (not shown), EOF reconstruction and optimal interpolation based on full covariance produced similar results, but these relations need further studies. When observed values were close to the model-computed climatological background, visual similarity was caused mainly by the dominance of spatial gradients of mean SSS over the spatio-temporal variability. Optimal interpolation with Gaussian approximation to the covariance produced realistic results in the neighbourhood of observation points, but gave unrealistic patterns and values in the distant SW extrapolation area.

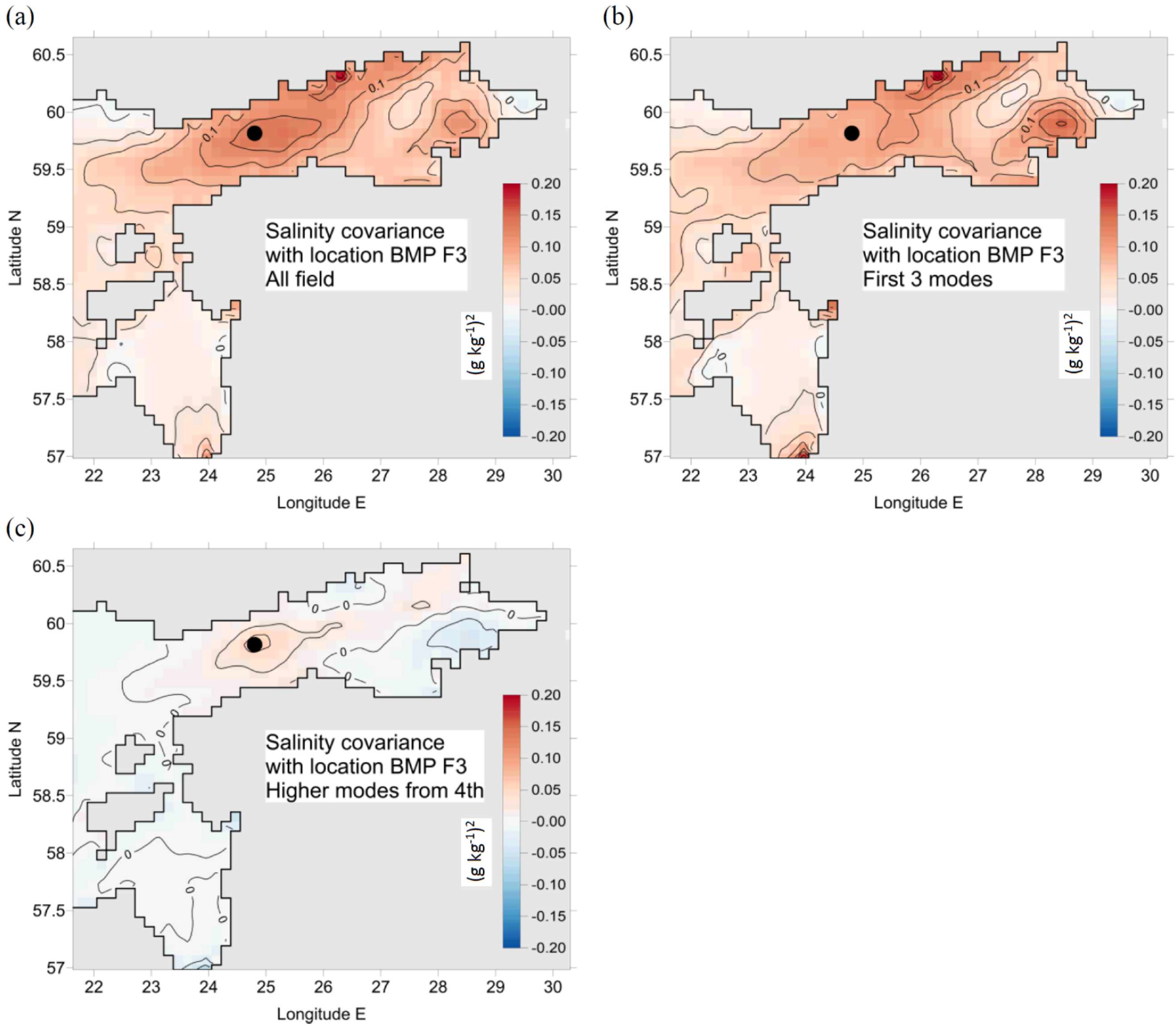

Figure x1: Spatial covariance of SSS with the values in grid cell near the HELCOM monitoring station BMP F3 (59.8383° N, 24.8383° E), extracted from the full covariance matrix calculated from the model data over 5 years. Covariance is decomposed by EOF modes: covariance of unfiltered data with all the modes included (a) is a sum of covariance of first three modes (b) and of the remaining higher modes, starting from the forth mode (c).

END

We copy here as an example one test figure (Fig. x2), that was not included in the revised MS, since it has not yet proved to be enough general.

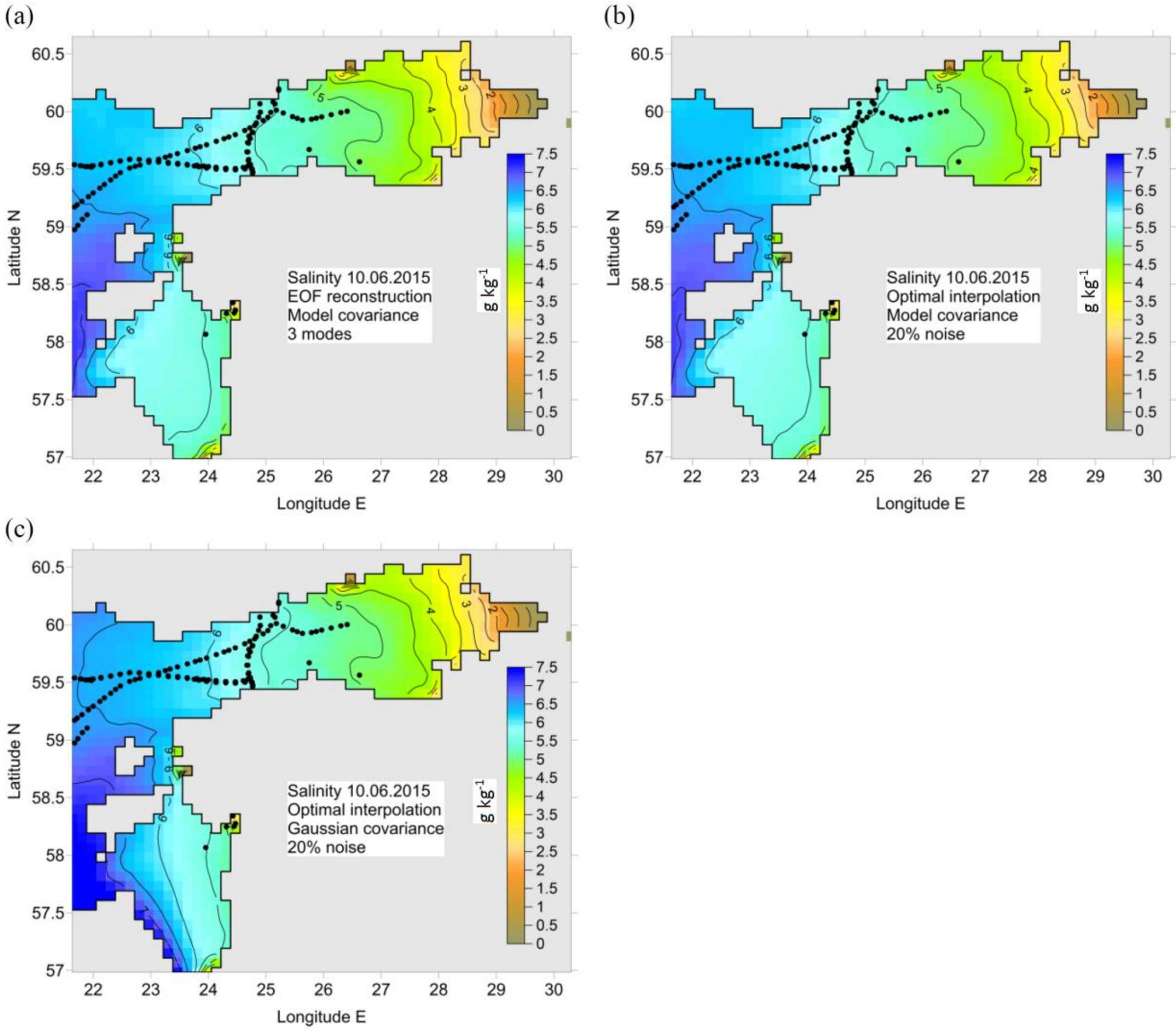

Figure x2: Example maps of reconstructing SSS based on full covariance matrix using EOF (a) and optimal interpolation (b), and optimal interpolation with Gaussian approximation of covariance, with spatial scale of 150 km.

- **We missed some discussion on the potential of the method to improve forecasts**

We have added following paragraph after line 487:
"We have tested the EOF-based DA in centred time window of 30 days, based mainly on available FerryBox data during the study period. As shown by reconstruction experiments by Elken et al. (2019), the time-dependent method can also work with backward observations as if it occurs during operational forecasts. When more observations become available, for example from new automated buoy stations, Argo floats and gliders, the time window can be shortened. Full covariance matrix estimated from the model results is the backbone of the EOF DA method. Prior and/or complementary to implementation of the method into operational practice, detailed covariance studies using results from multiple models could be useful, as well as additional reconstruction and DA studies using more data sources over longer periods."

- **We would furthermore appreciate some discussion about the implications of the assimilation on the model dynamics (e.g., vertical density structure, in particular stability)**

We have rewritten lines 497-499:
"There are obvious possible extensions of the EOF DA method to other variables and layers: improvement of stratification modelling, extension to biogeochemical models and DA of oxygen,

nitrogen and phosphorus. Applicability depends on how well the model reproduces the studied fields and their covariance, and much variance is explained by the major EOF modes."

- **It is not really clear, why the authors did not apply a more standard technique, with a more solid theoretical basis. A straightforward approach would be to use a low rank model error covariance matrix based on the presented EOF decomposition in the standard Kalman analysis equation. This would then also include observation errors and avoid the two steps required in the presented technique.**

We have added new section 3.1.1 Covariance, modes and reconstruction tests, given above. Using a full covariance matrix, optimal interpolation of the background field produced in several test similar results to the EOF reconstruction, but these relations need further studies.

Our results indicate that due to the imperfect observational network, model error covariance should also be treated by full covariance matrix. Approximated covariance was found to create too much distortion of the studied fields. Due to taking differences, the error covariance matrix could be more dependent on the model features than the background covariance matrix estimated from the validated model results. Because of absence of model error covariance estimates, we omitted the proposed option in the present study.

The presentation of the material should be improved. There are deficiencies, in particular with regard to putting the study into context of existing methods and motivating the selected approach. If "computational effort" is the main point, then this has to be quantified better.

Computational benefits are more elaborated. The paragraph on lines 231-235 is rewritten:

"The above DA method is computationally efficient. The EOF modes are calculated prior to DA cycles. For each DA time step, only one system of linear equations of rank of the number of EOF modes (about 3-6) has to be solved for the entire grid. The coefficients of the matrix are found by summation of the products of EOF mode values over the observation points (Eq. 2). For comparison, optimal interpolation requires solving the system of linear equations of rank of the number of observation points (about 100) for each grid cell (about 1000), with a single inverse matrix calculated for the time step."

There are quite a view grammar problems and a native speaker should proofread the text.

We plan additional language check.

We recommend publication after major revisions.

We have made substantial revision, added a new subsection and a new figure.

Specific Comments

Abstract

We think it would be better to structure the abstract such, that more general information (what is done?) comes first and specific results follow after that

When preparing the MS, the authors discussed both the options – your proposal and the one we have selected to present. Our choice is based on the better outreach possibilities, as we think.

Please explain acronym RMSD

Corrected

I think it is more common to say “dominating EOF modes” instead of “gravest”, but that should be checked by a native speaker

Corrected.

Introduction

Page 1, line 22: please reformulate “discrepancies of”

Changed to “errors of”.

Page 2, line 34: replace “then” by “the”

Corrected. Also, the first word of the sentence is replaced to “Whereas” (formerly “While”).

Data and methods

Page 3, line 72: “whichever” instead of “which”

Corrected.

Page 3, line 79: “… from the halocline …” please reformulate

The sentence has been reformulated: “... therefore <u>deeper</u> more saline waters ~~from the halocline~~ of the Baltic Proper penetrate into the Gulf of Finland and form an estuarine halocline ~~also there”~~

Page 3, line 87: “better grid cells” instead of “points”
It would be good to learn more about the vertical discretization of the model, e.g., how thick is the surface layer.

Corrected. The end of the sentence is modified “...71 986 of them on the surface <u>with a layer thickness of 3 m</u>”.

Page 4, line 95, maybe better “grid resolution” instead of “grid step”

Corrected.

Fig. 2a: Please change color of FerryBox tracks – it cannot be distinguished from land.
Please add information on the water depth the FerryBox observations are usually taken.

We have changed the color of land. We have added new sentences in the Sect 2.2, line 123:

“The analysed water is strongly mixed in the surface layer by the moving ship. Typical observation depth may be considered 5 m, although variations between the ships and due to the variable shipload exist (Lips et al., 2008; Karlson et al., 2016).”

Page 5: It was not clear, how you interpolate the FerryBox data to a 2D grid. Please explain in more detail.

The sentence was modified to:

“Two sets of compressed (averaged) FerryBox data were created for further data analysis, containing mean observed values, coordinates and observation times over the selected intervals.”

Page 5, line 142: Did you mean “… too irregular …” ?

Corrected.

Page 5, lines 146-150: this paragraph is hardly understandable – please reformulate.

The paragraph has been reformulated:

“The basic option of EOF reconstruction uses at each DA time step time-fixed amplitudes, encountering the observations spanning over certain time (which can be longer than DA time step) that are transferred to the fixed times by some interpolation or filtering/averaging procedure. The amplitudes are estimated ~~together with~~ using time-fixed observations by minimizing the root-mean-square-difference ~~(RMSD)~~ between the observations and the EOF reconstruction. The amplitudes at adjacent time moments are not directly related, but in case of longer temporal filters when observations overlap~~ping~~ ~~takes place~~ on different DA time steps, indirect relations between adjacent amplitudes become evident.”

Page 6, line 160: “the” instead of “then”

Corrected.

Section 2.3

The entire section is unfortunately quite messy and confusing, although (as far as I understand) the method is quite basic. The authors have to explain all symbols and indices with much more care. Also, what is a vector or matrix (what size?) and what is a scalar?

We have used the widespread notation that matrices and vectors are given in upright capital and lowercase bold letters, respectively, and scalars (including elements of matrices and vectors) are given in italic letters. There are two basic sizes of arrays, number of model grid points and number of observations. The presentation is a condensed version of subchapters “Notations for Empirical Orthogonal Functions (EOF)”, “Reconstruction of Observed Fields Using EOF Modes” and “Extension of the EOF Reconstruction Method to Time-Dependent Data” by Elken et al. (2019). Although the beginning of Section 2.3 says “...we chose to use EOF reconstruction of large-scale SST and SSS fields, using the orthogonal patterns from models following the detailed outline by Elken et al. (2019)”, we checked once more the clarity of condensed material and have rewritten the lines 156-159:

“The main steps of EOF reconstruction are the following. During the standard EOF decomposition, the orthonormal eigenvector matrix $\mathbf{E}$ (contains the spatial eigenvectors $\mathbf{e}_k$) is found from the eigenvalue problem $\mathbf{BE} = \mathbf{\Lambda E}$, where $\mathbf{B}$ is $M \times M$spatial covariance matrix, calculated from the $M \times N$ spatio-temporal matrix $\mathbf{X}$ of the “values of interest” by time averaging, and $\mathbf{\Lambda}$ is a diagonal matrix that contains eigenvalues λ_k.”

Page 6, line 185: “that that”

Corrected.

The basic assumption, if you use EOFs for interpolations like you do, is that the covariance structure of the model is correct – this should be stated more explicitly and discussed a little.

Problems of EOF reconstruction have been considered by Elken et al. (2019). We cite: SST and SSS results are rather well validated by observations and the model-based covariance patterns can be considered trustful. /// Fu et al. (2011) compared covariance patterns from modeled SST and satellite SST, and found them agreeing well. CMEMS QUID report has presented validation of SSS against FerryBox data, showing that the SSS patterns were well simulated by the model. In deeper layers, however, there is usually a larger spread between different model results.

Main differences between actual and model-based covariance estimates are expected within very short term variations (occurring above the Nyquist frequency/wavenumber) that comprise in observational datasets spatially uncorrelated noise, using the terminology of optimal interpolation.

We have added on line 166:
"If white noise with a variance ε^2 is present in the decomposed data due to sub-grid scale processes and/or sampling errors, the noise variance appears only as additive to the diagonal elements of the covariance matrix. The eigenvalue problem becomes $(\mathbf{B} + \varepsilon^2\mathbf{I})\mathbf{E} = \mathbf{\Lambda E}$, where $\mathbf{I}$ is a unity matrix. Patterns of spatial modes remain unaffected by adding the white noise, but the eigenvalues and energy share of the modes decrease according to a factor $(1 + \varepsilon^2/\sigma^2)^{-1}$. When the sum of eigenvalues of the included dominating modes is less than $\sigma^2 - \varepsilon^2$, contribution of noise is effectively smoothed."

You could have included observation errors in this interpolation exercise. I guess your assumption at the moment is, that the observations are 100% correct ? - please comment

Observation errors are considered in the revised text as follows.
Line 217
"This is the main DA calibration parameter, since extensive use of covariance statistics, including the effects of observation errors, has been included in the estimation of gridded reconstruction of point observations."

In the new sub-section 3.1.1
"EOF reconstruction method relies on the full covariance matrix, without any approximation. Full covariance matrix can be implemented in optimal interpolation as well. While EOF method needs to limit the number of included modes, smoothing in such way smaller scale variability and observational errors, optimal interpolation needs to include observational error variance ("nugget effect" in terms of Kriging method, equivalent to optimal interpolation); otherwise the system of underlying linear equations may become close to singular and the result may become unrealistically spiky."

Eq. 1: you assume that this matrix actually has an inverse – please comment. If the matrix is close to singular, you run into numerical problems as well.

Eigenvector matrix $\mathbf{E}$ is non-singular, since it is derived from the symmetric covariance matrix $\mathbf{B}$ on the basis of eigenvalue problem $\mathbf{BE} = \mathbf{\Lambda E}$. Inclusion of observation operator $\mathbf{H}_i$ (i is the assimilation time index) does not make the determinant of $\mathbf{E}^\mathbf{T}\mathbf{H}_i^\mathbf{T}\mathbf{H}_i\mathbf{E}$ equal to zero, if the number of observations is greater than zero. We excluded the situations with less than 6 observations and singularity was not detected. The cases with too large amplitudes were omitted and DA was not performed (see the text on lines 269-272).

Section 2.4

I guess eq. 1 is a continuous equation, which in its original form should be solved using the internal model time. I assume that you get eq. 4, if you replace the model time step by the assimilation time step – please explain more

Relaxation by Eq. (3) causes the model state to exponentially approach to the reconstructed grid (target) maps of observations ψ^o . If the restoring time scale τ is much longer than the model time step and still longer than the assimilation time step Δt, then it is sufficient to apply Eq. (4) with Δt.

I had problems to figure out how big the assimilation time step in the experiments actually is – please use consistent notation for critical parameters (e.g. time steps) throughout the document.

We admit that the notation Δt , with different indexes, has been used in the first version of the MS in too "distant" contexts - $\Delta t_p = t_p - t_i$ was the difference between the observation and reference times, Δ_R was the time window and Δt was the DA time step. We replaced the variation in time from $\Delta t_p = t_p - t_i$ to $\delta t_p = t_p - t_i$ and Δ_R to t_R.

The values of t_R and Δt were presented by words in the beginning of section 3.2: "...using the time-dependent EOF reconstruction method with a time window of 30 days..." and "Further on, each day DA was made on the fine grid using the procedure Eqs. (3)–(4)." We also added mathematical assignments.

Page 8: line 223: "The DA method … is analogical …" I don't think this is true in this generality, because it seems you don't consider observation errors at all. – please comment. The resemblance with 4DVAR is remote, because there is no model dynamics included in the minimization of the cost function.

The whole paragraph has been modified and unclear sentences were removed. The modified paragraph is:

"The DA method is based on the full covariance matrix of irregular pattern, calculated from model results over a sufficiently long period. Covariance is further treated using EOF modes. For the reconstruction procedure, we keep the lowest EOF modes without any approximation, covariance from higher modes is truncated. The large-scale features of the EOF reconstruction and associated DA exclude the possibility of creating spurious "bull-eye" patterns around observation points, that may happen for instance during unfavourable selection of optimal interpolation parameters. Subsequently, our DA method handles the large-scale features and excludes the possibility to assimilate smaller scale features, which can be described by the higher modes. The method of time-dependent amplitudes is able to encounter temporally distributed observations, when estimation of linear rate of change of the EOF amplitudes over the selected interval makes sense. Mesoscale deviations from basin-scale EOF patterns follow well-defined covariance decay with space lag; therefore, they could be treated by optimal interpolation with approximated covariance or similar methods (Elken et al., 2018)."

Page 8, line 239: "… artificial split …" I don't understand this sentence, because this "split" is a standard approach to validate assimilation techniques.

The sentence has been deleted in this section.

Page 7, line 217: " … since extensive use has been made …". This is I guess the critical point. The classical approach in an assimilation filter is to combine observations and the model state using covariance information on model errors at each analysis time step. In your approach there are no covariances of model errors. Instead, you use covariances of the background statistics for the interpolation. If you used a scaled version of the background covariance as a proxy for the model error covariance in a classical filter approach, you would probably end up with similar results, but with a more solid theoretical foundation. Anyway, as pointed out in the general comments, the method has to be put into the context of existing methods in a better way.

We used indeed the background covariance since validated model results are available. We have found that it has complicated structure, but can be physically well interpreted. Encountering the full covariance structure is very important, as we have shown, also in an example of optimal interpolation with full covariance structure. Covariance of model errors is not known in such details. We are not convinced that there is a simple transformation from background covariance to the model error covariance, since it has to be very model-specific, compared to the more universal estimates from validated model results.

We have added a new subsection 3.1.1 as pointed out earlier.

Section 3.1

Page 9, line 275: This is interesting; why don't you show the EOFS computed in your study?

These results were presented in detail by Elken et al. (2019). We found that repetition of figures is not necessary in this MS since there is open access to the earlier paper.

Page 13, line 408: The skill is often defined in relation to a reference run (e.g. the free run). In the case of the standard forecast skill, it is a dimensionless number – please check.

The paragraph has been rewritten:
"Ocean model performance (e.g. Stow et al., 2009; Golbeck et al., 2015; Placke et al., 2018) is usually evaluated by the differences between the observations and the model results, transferred to the times and locations of observations that they can be directly compared. The overall mean difference (over time and space) is termed bias and the standard deviation of differences at all the observation points is denoted as RMSD (centred root-mean-square difference). The forecast skill is usually non-dimensional, with the RMSD of the studied option (in our case, DA) scaled to reference data (FR in our case)."

Page 16: "There are obvious extensions … layers …"
This is, where it gets interesting, because the vertical structure of different model variables (temperature, salinity, etc) is a particular challenge and your assumption about the correctness of model covariances may become a problem (e.g., if the mixed layer thickness in the model is not correct)

We have extended the clause on line 498:

"Applicability depends on how well the model reproduces the studied fields and their covariance, and how much variance is explained by the major EOF modes."

Figure 8: It would be interesting to see the absolute differences between observations and the assimilation run and the same for the free run (these differences should reflect both observation and model errors).

This is a very interesting idea, but we think that adding more details to the figure will compromise readability too much.