Data assimilation of sea surface temperature and salinity using basin-scale EOF reconstruction: a feasibility study in the NE Baltic Sea

Mihhail Zujev[1], Jüri Elken[1], Priidik Lagemaa[1]

[1]Department of Marine Systems, Tallinn University of Technology, Tallinn, EE12618, Estonia

5   *Correspondence to*: Jüri Elken (juri.elken@taltech.ee)

Abstract. The tested data assimilation (DA) method based on EOF (Empirical Orthogonal Functions) reconstruction of observations decreased RMSD of surface temperature (SST) and salinity (SSS) in reference to observations in the NE Baltic Sea by 22% and 34%, respectively, compared to the control run without DA. The method is based on the covariance estimates from the long period model data. The amplitudes of the pre-calculated gravest EOF modes are estimated from point

10   observations using least-squares optimization; the method builds the variables on the regular grid. The study used FerryBox observations along four ship tracks from 1 May to 31 December 2015, and observations from research vessels. In the reconstruction, this data amount was compressed into daily averages over 5' N × 10' E coarse grid. Skill was tested based on daily averages on the 0.5' N × 1' E original fine grid of the model. DA with EOF reconstruction technique was found feasible for further implementation studies, since: 1) the method that works on the large-scale patterns (mesoscale features are neglected

15   by taking only the gravest EOF modes) improves the high-resolution model performance by comparable or even better degree than in the other published studies, 2) the method is computationally effective.

1 Introduction

In the coastal oceans and marginal seas, basin-scale observation, modelling and forecasting of oceanographic and biogeochemical variables is a continuing challenge. As an example from the Baltic Sea, large-scale nutrient dynamics

20   (Andersen et al., 2017; Savchuk, 2018) controls the level of eutrophication and hypoxia, affected by nutrient loads and changing climate (Meier et al., 2019). Placke et al. (2018) have recently shown, that Baltic Sea research models that do not include data assimilation, have still notable discrepancies of simulating sea surface salinity (SSS) and thermohaline stratification in the main basins, compared to the good accuracy of simulating sea surface temperature (SST). Similar evaluation has been obtained earlier by Golbeck et al. (2015), based on 13 operational models used routinely in the Baltic and

25   North Seas.

Data assimilation (DA) is a key element to improve the model accuracy with respect to observations, both in the operational forecast and the reanalysis context (Martin et al., 2015; Buiza et al., 2018; Moore et al., 2019). DA methods are built upon dynamical models and they are based on some kind of minimization (minimum variance, variational cost function formulation

30   etc.) of modelling errors (Carrassi et al., 2018), using estimated statistical characteristics of the studied variables. Most of the
widespread methods (optimal interpolation, 3DVar, 4DVar, various options of the Kalman filter, including their ensemble
formulations) use covariance as the basic statistical characteristic. Recent overviews on different DA applications in the Baltic
Sea can be found in the papers by Liu and Fu (2018), Zujev and Elken (2018), Goodliff et al. (2019), She et al. (2020). While
there are several results from Baltic Sea reanalysis studies available (Axell and Liu, 2016; Liu et al., 2017), then operational
35   forecasts within CMEMS (Copernicus Marine Environment Monitoring Service) do not include presently DA (Huess, 2020).

Results of DA based forecasting depend heavily on the spatio-temporal configuration of the observing system (LeTraon et al.,
2019). Unlike the regular weather observing networks, observation systems in marginal seas are rather fragmented, where
areas and periods of dense sampling can be neighboured by large observation gaps. Therefore, special OSE (Observing System
40   Experiment) studies have been initiated, to find optimal observation network configurations to achieve best skill of DA (Fuji
et al., 2019). However, most of the observations of the Baltic Sea surface variables, not yet detectable by remote sensing (like
salinity, nutrients etc.), stem from the FerryBox systems installed on board regularly cruising commercial passenger or cargo
ships (She, 2018), and planning can be done only within the existing routes. Therefore, development of improved gap-filling
techniques is a challenge.

45

Recently, a novel method for EOF reconstruction of gridded SST and SSS fields, using the data from (mostly) irregular and
(often) sparse observations was presented as an idea (Elken et al., 2018), and then it was thoroughly tested in the NE Baltic
Sea (Elken et al., 2019). The method relies on the estimate of covariance matrix from the long-period model data, which is
decomposed into the full set of EOF modes. The mode values at observation points, together with the observed values, enable
50   least-squares estimation of observational amplitudes. The method is able to follow on the regular grid the pointwise observed
temporal changes of the mean state and of the major basin-scale gradients. The aim of the present study is to implement this
statistical reconstruction technique into the data assimilation of the forecast model, and to study the feasibility of such
assimilation method.

55   The paper is organized as follows. In the section of data and methods, firstly, an overview of sub-regional oceanographic
background and short model description are presented. Observational in situ data have been compiled from three sources, and
they contain shipborne monitoring and FerryBox platforms. The reconstruction method is presented in detail, and the section
ends with the description of the used data assimilation method. The results section starts with the presentation of experiments
in order to find the optimized parameters for reconstruction of gridded fields. The rest of the section is devoted to the analysis
60   of the results of data assimilation experiments, ending with the skill evaluation. Finally, discussion and conclusions are
presented.

2 Data and methods

2.1 Study area and the circulation model

We have chosen the study area in the NE Baltic within 56.983°–60.65° N, 21.633°–30.3° E (Fig. 1), motivated by several
65 Estonian national interests within the operational forecast of sea state and assessments of the marine environment. The region
covers the Gulf of Finland, the Gulf of Riga and part of the Baltic Proper adjacent to these gulfs. The region is rather shallow:
the mean and maximum depths are 26 m and 62 m in the Gulf of Riga (Yurkovskis et al., 1993) and 37 m and 123 m in the
Gulf of Finland (Alenius et al., 1998), respectively.

70 The region lies in the temperate climatic zone. During the summer, yearly maximum SST exceeds usually 15 °C in July or
August (Alenius et al., 1998), with highest values up to 25 °C observed in some years in the shallow coastal zones (Stramska
and Białogrodzka, 2015). The warm upper layer of 10–20 m thickness is well mixed down to the thermocline or bottom, which
of them is shallower. Occasionally, wind-driven coastal upwelling processes disrupt this warm layer (Uiboupin and Laanemets,
2009). Nearly every winter, sea ice forms with variable extent and thickness; during severe winters, the Gulf of Finland and
75 the Gulf of Riga are fully ice-covered (Jevrejeva et al., 2004). The region is impacted by large rivers: the Gulf of Finland and
the Gulf of Riga together receive 34% of the total freshwater discharge to the Baltic Sea as can be calculated from the data by
Johansson (2017). As a result, there is estuarine increase of SSS from east to west (Alenius et al, 1998; Yurkovskis et al.,
1993), reaching 7–8 g kg^{-1} in the Baltic Proper (Kõuts and Omstedt, 1993). The Gulf of Finland has a free connection to the
Baltic Proper without sill or any other topographic restriction, therefore more saline waters from the halocline of the Baltic
80 Proper penetrate into the Gulf of Finland and form an estuarine halocline also there (Liblik et al., 2013). A shallow sill of the
depth of 15 m connects the Gulf of Riga with the Baltic Proper; therefore deep layers of the Gulf of Riga can receive only
surface waters of the Baltic Proper (Lilover et al., 1998). The two gulfs, located in the NE Baltic, play an essential role to the
dynamics of the whole Baltic Sea (Omstedt and Axell, 2003).

85 For the modelling, Estonian sub-regional setup (Fig. 1) of the Baltic-wide HBM model was applied with 0.5' N × 1' E
resolution containing the entire Gulf of Finland, Gulf of Riga and NE portion of Baltic Proper (Lagemaa, 2012; Zujev and
Elken, 2018). The model fields are three-dimensional having 455 × 529 × 30 points (by latitude, longitude and depth
correspondingly) with 750 088 wet-points, and 71 986 of them on the surface. At the western open boundary, the data were
taken from the Baltic-wide HBM model, operated by the Copernicus Marine Environment Monitoring Service (CMEMS,
90 https://marine.copernicus.eu/). Atmospheric forcing was provided by the Estonian implementation of HIRLAM. HBM uses
the Arakawa C-grid, and produces forecast for 16 ocean variables including temperature, salinity, current speed and ice
concentration. Detailed description of the model and its validation can be found by Berg and Poulsen (2012).

The model setup has been designed for operational forecast. For computational reasons, it was decided to keep the operational
95   0.5 nautical mile grid step and to perform shorter feasibility experiments, instead of choosing larger grid step and making
longer experiments. The model is used routinely by the Estonian Weather Service (implemented by one of the authors, Priidik
Lagemaa); SST is displayed on the web page https://ilmateenistus.ee/meri/mereprognoosid/merevee-temperatuur/ and SSS is
shown on the page https://ilmateenistus.ee/meri/mereprognoosid/soolsus/. In compliance with the recent study by Zujev and
Elken (2018), we chose the study period from 1 May to 31 December 2015, to be used for the DA experiments. The model
100   experiments were conducted in the framework of operational forecast, where the forcing files were downloaded daily. There
were no gaps during the study period in meteodata nor in open boundary conditions nor any other input.

2.2 Observational data

All available SST and SSS data from three sources were compiled:

1.   Copernicus Marine Environment Monitoring Service (CMEMS, https://marine.copernicus.eu/) contains among other
105   data sources the quality-checked data set of Baltic in-situ near-real-time multiparameter observations ftp://nrt.cmems-
du.eu/Core/INSITU_BAL_NRT_OBSERVATIONS_013_032/bal_multiparameter_nrt, downloaded on 24 October
2019. This data set, accessible through free of charge registration, contains in our study region data from several
FerryBox systems (automatic observations made from ferries and other ships crossing the sea areas on a regular basis).
There is also a number of coastal stations, but they record mainly sea level and water temperature, whereas salinity
110   observations are missing; therefore we are not using coastal stations. In our study area and time interval, there were
not any operating buoy stations, gliders or Argo floats.

2.   HELCOM/ICES database contains the results from the HELCOM marine monitoring programme and is hosted by
ICES (https://ocean.ices.dk/helcom, data downloaded on 22 October 2019). It includes mainly the data from
shipborne monitoring stations, where SST and SSS are easily extracted.

115   3.   National monitoring database KESE (https://kese.envir.ee/kese/listProgram.action, search for "mereseire"), contains
detailed records of all variables observed under the national environmental monitoring program. The data that were
downloaded on 18 October 2019, contain different data records for every environmental variable. Except for a few
cases, these data are also found in the ICES/HELCOM database. Duplicate entries were avoided from the composite
data set by averaging over small time and space intervals.

120

The largest amount of synchronous SST and SSS data originates from the FerryBox systems, accessed through the CMEMS
(Table 1). There were about 330 k (thousand) initial observation points from FerryBox, distributed over a few ship lanes
(Fig. 2a) with a few hundred meters resolution and from daily to a few days interval. There were also about 370 observations
from shipborne monitoring stations. Distribution of the amounts of observations in selected temporal and longitude intervals
125   (Fig. 2b) reveals a highly irregular pattern. Most of the observations were concentrated on the Tallinn-Helsinki transect
located across the Gulf of Finland between the longitudes 24.6°–25° E. FerryBox observations were missing in the Gulf of

Riga and in the eastern part of the Gulf of Finland, east from 26.5° E. In the southern part of the Gulf of Riga, available data were missing during the study period.

130 Two sets of compressed FerryBox data were created for further data analysis. Firstly, for the validation and skill study, daily mean spatial averages over a fine grid (0.5' N × 1' E as in the used model) cells were created, resulting in about 110 k values. Secondly, for the EOF pattern analysis and reconstruction of SST and SSS fields, daily mean spatial averages over the coarse grid (5' N × 10' E) were created. In this procedure, the initial observations were compressed on the coarse grid by roughly 25 times yielding about 13 k average values for SST and SSS. Within the temporal averaging, it was not chosen to

135 apply any diurnal cycle correction and all the observations at different hours were averaged to the closest midnight. For the interpretation of model and DA results, meteorological data were taken from the model forcing fields. For the occasional comparison, CMEMS remote sensing SST Level 4 (L4) data were retrieved from the service portfolio http://marine.copernicus.eu/services-portfolio/access-to-products/ as the product SST_BAL_SST_L4_NRT_OBSERVATIONS_010_007_b.

140 **2.3 Reconstruction of gridded data from point observations**

For the purpose of DA, 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). The spatio-temporal distribution of data was irregular to use standard interpolation and filtering algorithms like the Cressman method or optimal interpolation (see an example from the same region by Zujev and Elken, 2018).

145

The basic option of EOF reconstruction uses time-fixed amplitudes, encountering the observations that are transferred to the fixed time by some filtering/averaging procedure. The amplitudes are estimated together with 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 overlapping takes place, indirect

150 relations between adjacent amplitudes become evident.

Elken at al. (2019) proposed also an advanced method with time-dependent amplitudes. Within this approach, the amplitudes and their time derivatives are estimated together with observations within a selected time interval, in order to minimize RMSD between the observations and EOF reconstruction in the observational framework.

155

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 a covariance matrix, calculated from the 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. The dataset $\mathbf{X}$ contains time slices $\mathbf{x}_i$ that are spatial state vectors at time i. While $\mathbf{E}$ is non-

160   dimensional, then dimensional amplitudes (or in other words, factors) of EOF decomposition are found by $\tilde{\mathbf{a}}_i = \mathbf{E}^{\mathrm{T}} \mathbf{x}_i$, and the decomposition is reconstructed to the "values of interest" by $\mathbf{x}_i = \mathbf{E} \tilde{\mathbf{a}}_i$. Here we have used the notation $\tilde{\mathbf{a}}_i = \mathbf{\Lambda} \mathbf{a}_i$, where $\mathbf{a}_i$ is non-dimensional amplitude. The eigenvalues λ_k present the variance (energy) of the eigenvectors $\mathbf{e}_k$ over the whole period, the sum of all eigenvalues equals to the variance of $\mathbf{X}$. EOF decomposition offers the possibility to keep only the most energetic modes in the reconstruction and truncate the higher modes in $\mathbf{E}$. When L most energetic modes are taken into account in the

165   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.

During EOF reconstruction from observations $\mathbf{y}_i$, the number of observations K is assumedly smaller than the number of points M in the spatial eigenvectors $\mathbf{e}_k$ that are determined on the model grid and evaluated from the model statistics. For the

170   comparison with observations, the model data $\mathbf{x}_i$ are transformed to the observation points by the observation matrix $\mathbf{H}_i$ by the formula $\mathbf{H}_i \hat{\mathbf{x}}_i = \mathbf{H}_i \mathbf{E} \hat{\mathbf{a}}_i$, where $\hat{\mathbf{a}}_i$ are the "observational" amplitudes. Further, the $\hat{\mathbf{a}}_i$ values should follow least-square minimization of reconstruction error in relation to observations $\|\mathbf{y}_i - \mathbf{H}_i \mathbf{E} \hat{\mathbf{a}}_i\|^2 \Rightarrow \min$. The expressions to find observational amplitudes and reconstructed fields are

175    $$\hat{\mathbf{a}}_i = \left(\mathbf{E}^{\mathrm{T}} \mathbf{H}_i^{\mathrm{T}} \mathbf{H}_i \mathbf{E} \right)^{-1} \mathbf{E}^{\mathrm{T}} \mathbf{H}_i^{\mathrm{T}} \mathbf{y}_i \, , \qquad \hat{\mathbf{x}}_i = \mathbf{E} \hat{\mathbf{a}}_i . \tag{1}$$

In the reconstruction by Eq. (1), the critical point is a possibility of spurious amplitudes based on few and unfavourable spaced observation points. Experiments with pseudo-observations (Elken et al., 2019) revealed that the values of $\hat{\mathbf{a}}_i$ should match the limits derived from statistics of $\tilde{\mathbf{a}}_i$, whereas higher modes with outlying amplitudes should be neglected. In practice, the cycle

180   of truncation of $\mathbf{E}$ starts from the accounting of lower, most energetic modes, and continues up to L modes until the values of $\hat{\mathbf{a}}_i$ are in acceptable limits.

Most of the oceanographic observations are not instant in time. It may take several days or even weeks to cover larger sea area with shipborne monitoring. When P observations $\mathbf{y}_p$ are taken at different times p, then construct an observation matrix $\hat{\mathbf{H}}_p$

185   that that allows pointwise comparison of $\mathbf{y}_p$ and $\hat{\mathbf{H}}_p \mathbf{x}_i$ converted from gridded values at specified time i. Assume that within the short time span the amplitudes depend linearly on time and introduce $\hat{\mathbf{b}}_p = \hat{\mathbf{a}}_i + \mathbf{d}_i \cdot \Delta t_p$, where $\hat{\mathbf{a}}_i$ is the time-fixed amplitude, $\mathbf{d}_i$ is the rate of change vector and $\Delta t_p = t_p - t_i$ is the difference between the observation and reference times. The function to be minimized regarding reconstruction errors is $Q = \left\| \mathbf{y}_p - \hat{\mathbf{H}}_p \mathbf{E} \hat{\mathbf{b}}_p \right\|^2 = \left\| \mathbf{y}_p - \hat{\mathbf{H}}_p \mathbf{E} \left(\hat{\mathbf{a}}_i + \mathbf{d}_i \cdot \Delta t_p \right) \right\|^2$, which yields a system of $2L$ linear equations obtained from $\partial Q / \partial \hat{a}_l = 0$, $\partial Q / \partial d_l = 0$, $l = 1 \dots L$

190

$$\mathbf{G} \mathbf{z} = \mathbf{w} \, , \qquad G_{mn} = \sum_{p=1}^{P} f_m^p f_n^p, \qquad w_n = \sum_{p=1}^{P} y_p f_n^p . \tag{2}$$

Here the vector of unknowns combines the amplitudes and their rates of change $\mathbf{z} = \{\hat{a}_1 \dots \hat{a}_L, d_1 \dots d_L\}$. Instead of the full set of EOF mode values, as during standard decomposition, we take the modified/interpolated mode values at observation points;

195   then $f_m^p = \{\hat{e}_1^p \dots \hat{e}_L^p, \hat{e}_1^p \Delta t_p \dots \hat{e}_L^p \Delta t_p\}$. We note that when all observations have the same time stamp and $\Delta t_p = 0$, the Eq. (2) is reduced to (1).

Time-dependent reconstruction allows selecting the reference time and length of time interval. As with the time-fixed reconstruction, the highest "usable" mode is determined by checking the amplitude values with statistical limits. The method

200   also allows estimation of amplitudes and making reconstruction by only backward observational data. This feature makes the method useful in operational forecasts, where only past observations can be taken into account for drawing the present nowcast maps.

2.4 Method for data assimilation

Many DA techniques use (irregular) point observations of a variable ψ as the input source. In our approach, gridded maps ψ^o

205   are used; they are optimized by EOF reconstruction as described in Sect. 2.3. Therefore, in the continuous equivalent, DA is performed by Newtonian relaxation (e.g. Holland and Malanotte-Rizzoli, 1989)

$$\partial\psi/\partial t = F(\psi) - \frac{1}{\tau}(\psi - \psi^o) , \qquad\qquad (3)$$

210   which discrete form has been applied for DA using gridded climate data (Moore and Reason, 1993) or using daily satellite-based SST data (Ravichandran et al., 2013). Equation (3) is then written for DA time step Δt in two stages as

$$\psi^f = \psi^{a-1} + \Delta t\, F(\psi^{a-1}) , \quad \psi^a = (1-\alpha)\psi^f + \alpha\psi^o, \qquad\qquad (4)$$

215   where ψ^f is the raw forecast field calculated from the previous analysis field ψ^{a-1} without DA using the model operator F and ψ^a is the new analysis field. Equation (3) contains adjustable relaxation time τ that is transformed in Eq. (4) to non-dimensional $\alpha = \Delta t/\tau$. This is the main DA calibration parameter, since extensive use of covariance statistics has been included in the estimation of gridded reconstruction of point observations.

220   In practical calculations, SST and SSS observational data were reconstructed on the coarser 5' N × 10' E grid and interpolated/extrapolated by bilinear procedure to the finer 0.5' N × 1' E model grid. This approach is justified by irregular distribution of observations; reliable estimation is possible for only large-scale patterns of SST and SSS fields.

The DA method based on the time-fixed amplitudes of reconstruction is analogical to optimal interpolation (OI) and 3DVar methods (in this study we have truncated the vertical coordinate, but three dimensions can be easily included in the procedure),

225 and the method of time-dependent amplitudes is similar to 4DVar method. The basic difference appears in the handling of the covariance - we use the model results over a longer period to pre-calculate full covariance matrix of irregular pattern and keep the lowest EOF modes without any approximation, covariance from higher modes is truncated. Subsequently, our DA method handles the large-scale features and excludes the possibility to assimilate mesoscale features, which can be described by the higher modes. Mesoscale deviations from basin-scale EOF patterns follow well-defined covariance decay with space lag;

230 therefore, they could be treated by optimal interpolation or similar methods (Elken et al., 2018).

The above DA method is computationally efficient. For each time step, only one system of linear equations has to be solved for the entire grid. For comparison, optimal interpolation requires solving the system of linear equations for each grid cell. The rank of the system matrix of the time-dependent EOF method is two times the number of the used gravest EOF modes. The

235 coefficients of the matrix are found by summation over all the grid cells.

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. Therefore, all the observational data were included in DA and artificial split between "assimilated" and

240 "independent" observations was not made.

The model skill with respect to observations was evaluated over those grid cells - time span pairs when observations were available. Since observations covered only a small part of the study domain, DA results were also compared with control run without DA, but then it is possible to only analyse the changes due to DA, without conclusion of possible improvement.

245 Standard statistical characteristics were calculated for individual fields: mean, standard deviation, in case of differences (for example, relative to observations): bias, RMSD (root-mean-square difference, equals to the standard deviation of difference field), and the Pearson correlation.

3 Results

3.1 Finding the parameters for reconstruction of gridded fields

250 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). Multiple checks performed on our data set suggested that three gravest modes were included in the EOF reconstruction. Reconstruction experiments were made

with different intervals (time window) Δ_R around the reference time t_i; including the observations within time window from $t_i - \Delta_R/2$ to $t_i + \Delta_R/2$. The results were evaluated to fulfil the goals:

255   A. Small RMSD between the observed values and the reconstructed fields;

         B. Small number of gaps in the reconstructed time series;

         C. Low number or missing presence of "spikes" and/or "jumps" in the time series.

Two basic options for temporal handling of the reconstruction procedures were tested:

260 (a) application of procedure by Eq. (1) of time-fixed amplitudes; time average of observations was taken for each grid cell, time adopted in each grid cell as constant reference time,

(b) full application of the procedure by Eq. (2) of time-dependent amplitudes; all the daily mean observations (average was taken also over coordinates and time) were kept separate for each coarse grid cell where the observations existed.

265 In addition, procedure by Eq. (2) was tested with an option with time average of observations in each grid cell, and with selection of closest to the reference time observations. These experiments provided more spikes and 70% higher RMSD than the basic options (a) and (b) and they were neglected from further consideration.

As a first step in all the experiments with variable time window, the EOF amplitudes of the mode k were checked for the limit

270 $|\hat{a}_{i,k}| < 2\sqrt{\lambda_k} = 2\sigma(\tilde{a}_k)$, where σ denotes standard deviation. DA data for the days with higher amplitudes were left blank since these reconstruction results most frequently got "out of reality". In addition, when the number of observations was less than six, reconstruction was not performed.

Time window for the experiments (a) and (b) were selected 10, 20 and 30 days. Elken et al. (2019) have found that the

275 correlation time scales (e-fold drop, correlation value 0.368) of EOF SST amplitudes were 65 days for the seasonal 1st (overall heating/cooling) and 2nd (faster heating/cooling in shallow coastal areas) modes, and 15 days for the 3rd "upwelling" mode. Time scales of the SSS modes were 65 days for the 2nd and 3rd mode, representing the large-scale gradients, and 110 days for the 1st mode describing long-term variations of mean salinity.

280 Methods of time-fixed (a) and time-dependent (b) reconstructions revealed similar statistical results during the study period in 2015, whereas RMSD between observed and reconstructed values of (a) was by 5% larger than of (b). By increasing the time window, RMSD of reconstruction slightly increases due to the stronger smoothing. The smoothing effect can be seen from the reconstruction examples given in Fig. 3. It should be noted that the reconstruction is designed to yield the best approximation to the observations over the entire region; therefore, it does not need to present the local best fit at individual points.

285

Network of observations, available during the study period, appeared favourable for the reconstruction, although observations were missing in the southern part of the Gulf of Riga and eastern part of the Gulf of Finland. With a time window of 30 days, there were no reconstruction gaps identified during the study period, determined for both of the methods by the above described amplitude limit criteria. Smaller time windows yielded some gaps in 2015. During the longer period from 2010 - 2018, gaps were found in most of the years (except our study period), whereas shorter time windows result in more reconstruction gaps. Detailed comparison of the time-fixed (a) and time-dependent (b) methods revealed that time-fixed reconstruction might create spurious "jumps" when there is a gap in observations which length is close to the time window. In that case, backward average is taken before the gap and forward average after the gap, which may result in "jumpy" results. Time-dependent reconstruction, which also accounts for the temporal changes within the time window, handled such situations more smoothly.

Based on the results from reconstruction experiments, the gridded SST and SSS data were reconstructed by the time-dependent reconstruction method (b) using three gravest modes in the time window of 30 days, centred around the assimilation time like during reanalysis. These gridded fields were applied in the DA relaxation scheme Eqs. (3)–(4).

3.2 Data assimilation experiments

In DA experiments, gridded observational data were pre-calculated each day using the time-dependent EOF reconstruction method with a time window of 30 days as presented in Sect. 3.1. Reconstructed SST and SSS fields were calculated on the coarser five nautical miles grid and interpolated bilinearly to the fine 0.5 nautical mile grid. Further on, each day DA was made on the fine grid using the procedure Eqs. (3)–(4). Two basic experiments were conducted, with relaxation time 10 days (weight of observations 0.1, experiment code DA01, assimilated fields SST01 and SSS01 for temperature and salinity, respectively) and with relaxation time 5 days (weight 0.2, experiment code DA02, field codes SST02 and SSS02). In addition, a variety of short-term trials was performed in a preparatory phase (results graphically not presented) which led to the two basic experiments. Comparison data were coded as FRT and FRS for temperature and salinity of control run without DA, and FBT and FBS for observed FerryBox data, respectively.

3.2.1 Example from the beginning of August

There was an interesting oceanographic situation in the beginning of August, when moderate but extensive upwelling SST pattern at the northern coasts of the basins (Fig. 4), with some effects on SSS (Fig. 5), was combined with fast heating of a thin (6-9 m) surface layer (Fig. 6). Since the middle of July, moderate winds with speeds from 4 to 6 m/s, which had a westerly zonal component (favoring upwelling at the northern coasts of the basins), were blowing above the Gulf of Finland. After 3 August 2015 (the maps in Figs. 4 and 5 are taken on this date), wind ceased and air temperatures increased by 10 August across the study area up to 25–27 °C in the Gulf of Finland and up to 31 °C in the southern Gulf of Riga, creating a thin layer of warm surface water. Heating of surface waters was favored by high nightly air temperatures, higher than SST. Vertical profiles (not shown) in the Gulf of Finland revealed deep thermocline at 40 m depth near the southern (downwelling) coast and shallower

thermocline near the northern coast; the warm surface water column was near Tallinn two-three times thicker than near Helsinki. From the end of July to 10 August, warming resulted in an increase of SST (Fig. 6) near Tallinn from 16.5 °C to 18.5

320 °C and near Helsinki from 14.5 °C to 18 °C.

The SST maps presented in Fig. 4 include control run, satellite observations and two experiments with DA. When warm waters with SST above 17 °C dominated the study area, all the maps revealed moderate upwelling near the northern coasts of the basins. However, the minimum temperatures and the spatial extent of the colder waters were different. Warmest "cold" waters

325 were observed on satellite images. While satellites measure SST of a thin surface layer, then FerryBox and models acquire temperature over much thicker layer. It is known that in the Gulf of Finland satellite and FerryBox can have similar SST values in case of winds stronger than 5 m/s (Uiboupin and Laanemets, 2015); at smaller wind speeds the SST bias can be 1–3 °C in reference to FerryBox observations. Within these accuracy limitations, satellite observations presented in Fig. 4d confirm the model patterns to some extent. The control run (Fig. 4a) was characterized by too high SST contrasts, compared to the satellite

330 data (Fig. 4d). From the earlier study by Zujev and Elken (2018), it is known that the free model without DA forecasts faster heating and cooling of shallow coastal areas and slower heat dynamics in offshore areas. Data assimilation, made using the FerryBox data (Fig. 4b, c) reduced discrepancies with satellite observation. The major large-scale differences between the satellite data (Fig. 4d) and the best DA02 (Fig. 4c) can be outlined as follows: (1) the colder upwelling water extended on the satellite image further to the east, (2) warmer waters were found on the satellite images in the southern Gulf of Riga, near the

335 Daugava river, and in the shallow areas between the Estonian islands, (3) in the Gulf of Riga, a strip of colder waters was modelled along the western coast, while satellite observations revealed warmer waters near this coast.

There were also numerous mesoscale features evident on SST (Fig. 4) and SSS (Fig. 5) maps, like colder upwelling filaments along the northern coasts of the Gulf of Finland and the Gulf of Riga, and decaying anticyclonic warm-core eddies near the

340 southern coast of the Gulf of Finland. The Irbe Front (Lilover et al., 1998; Raudsepp and Elken, 1999), formed by the salinity difference between the Gulf of Riga and the Baltic Proper, was found by the SSS maps in the outward position, stretching from the strait towards the open sea. This salinity structure was also repeated in the SST patterns; the satellite observations confirmed the predicted outward position during the taken snapshot. The model predicted that in the Gulf of Riga the Daugava river waters were spreading by narrow coastal strips of smaller salinity in both the NE and NW directions (Fig. 5).

345 **3.2.2 Time series in the areas of dense observations**

Locations with dense observations allow us to validate the model and visually evaluate assimilation quality. We compared SST and SSS data of control run (FR) and DA options DA01 and DA02 with FerryBox data (FB) at two points near Tallinn and Helsinki (Fig. 6). While SST followed the seasonal cycle, with weather-dependent deviations, then SSS behaviour was more irregular. In the given variation scales of SST and SSS (16 °C and 2 g kg^{-1} respectively), all the compared SST data

350   sources were more similar to each other than that of SSS. Still, most of the time the assimilation curve (DA02) was closer to
       the FerryBox observations than the control run, for both SST and SSS.

       Warm conditions in the beginning of August (Sect. 3.2.1) are clearly visible on SST time series (Fig. 6a, c). Comparing the
       values near Tallinn and Helsinki, the southern part of the Gulf of Finland was roughly 2 °C warmer than the northern part,
355   whereas the northern part had an unstable day-to-day pattern, possibly due to the fluctuations of the upwelling pattern. This is
       consistent with the spatial maps given in Fig. 5. Near the southern coast, an upwelling event occurred in September, reducing
       SST during a few days nearly by 4 °C (Fig. 6a). Larger SST drop during the southern coast upwelling (at easterly winds),
       compared to the northern coast upwelling (at westerly winds of the same magnitude), is explained by the steeper topography
       slopes in the southern part of the Gulf of Finland (Laanemets at al., 2009). This upwelling event was properly resolved by all
360   the data sets, with DA02 being closest to observations. In general, a free model without DA expected warming at a lower rate
       during summer and was more precise in autumn, while both assimilation experiments properly corrected the SST and SSS
       values. However, in some cases, assimilated temperature was somewhat higher than observed and modelled SST.

       Assimilation resulted in one major SSS improvement in early summer when the model predicted upwelling with too high
365   salinity near Helsinki. Nevertheless, in some cases DA made minor corrections at one of the locations, ignoring observations
       and sticking to the control run (e.g. late July - early August near Tallinn, October near Helsinki). When the model overshoots
       at both locations, DA properly corrects temperature and salinity values. This implies that DA tends to better correct the mean
       values than the cross-gulf gradients.

       In the salinity time series, a "freshwater event" with reduced salinity was observed in the Gulf of Finland at the end of
370   September and beginning of October. In the daily SSS data (Fig.6b, d) the event was spiky, possibly due to the mesoscale
       features not assimilated in the present study: without DA, the eddies tend to have random phase, and the spikes in the time
       series of different model options and observations do not need to be coherent. However, in the weekly averaged data (not
       shown) the mesoscale activity was suppressed and the fresh event appeared simultaneously in all the data within the central
375   and western part of the Gulf of Finland.

       Regarding the assimilation experiments DA01 and DA02, there is no proportionality between the options of 5 and 10 relaxation
       days in terms of DA performance, as can be seen from Fig. 6. They diverged as the region experiences a temperature drop or
       daily trend change. Both options of assimilated SST could either coincide for a long time or go in parallel, but DA02 was
380   systematically closer to the FerryBox observations. Salinity fluctuations had larger amplitudes in the free run without
       assimilation, but both DA algorithms, with a "thumb" rule - the bigger the weight, the bigger the change, had properly corrected
       them. Only in December DA01 showed better results, being closer to the FerryBox salinity than assimilation DA02.

3.2.3 Spatio-temporal dynamics

We have chosen to compare assimilation with best results (DA02) to the control run without data assimilation (FR), and track
385 the continuous time-latitude changes of SST and SSS (Fig. 7) in two sub-basins - Gulf of Finland and Gulf of Riga along the
coast-to-coast transects given in Fig. 2a. Using DA, temperature was corrected approximately by 1–2 °C, and salinity by less
than 1 g kg^{-1}. Major systematic change (in the Gulf of Finland this was validated as improvement, see further Sect 3.2.4) was
seen near the coasts and in spring/autumn periods, while summer temperatures underwent minor corrections. Salinity
corrections had more uniform distribution and smooth drifting pattern - DA consistently increased SSS values with time and
390 southwards in both of the sub-basins.

Data assimilation had increased SST in the Gulf of Finland in open waters during the warming period and in late autumn all
across the gulf, and had decreased in the coastal areas during the warming period, whereas near the northern coast this decrease
continued until September. In the Gulf of Riga, SST increase dominated throughout the study period, but it was interrupted
395 occasionally by basin-wide events when DA had decreased the FR temperature. Largest corrections of both SST and SSS were
evident in the coastal waters. Salinity was increased by DA in most of the cases in the Gulf of Finland, except for May-July
near Tallinn. Largest increase of SSS occurred in November and December, when control run results dropped compared to the
earlier period.

400 Some unusual basin-wide events can be found on the difference charts in Fig. 7. For example, abrupt warming of the surface
around 10 August 2015 (Sect. 3.2.1) was correctly predicted by the free run model (Fig. 6c), but it was over-smoothed by the
data assimilation. Similar line in December on both charts denotes occurrence of fronts of cold and saline water due to strong
winds and storms.

405 As there are not enough observations available in the Gulf of Riga for validation, we cannot definitely say whether DA
improved the situation in the region and to what extent.

3.2.4 Evaluated skill

Ocean model performance or skill (e.g. Stow et al., 2009; Golberg 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
410 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 centred RMSD (root-mean-square deviation).

The present ocean model has a fine 0.5' N × 1' E resolution of about 930 m (Sect. 2.1), therefore for comparison with
observations we used a simplified approach and took averages of observations over the model grid cells over daily time span

415   (Sect. 2.2). Such compressed fine-resolution observational data set, still having about 110 k points for SST and SSS, was
originating mainly from the FerryBox (FB) lines (Fig. 2), and it covered central and western parts of the Gulf of Finland and
the neighbouring part of the Baltic Proper. Areas with lower salinity in the eastern Gulf of Finland and in the Gulf of Riga did
not have any massive observations.

420   Data from the DA experiments DA01 and DA02 were compared to the same compressed observational FB data as the data
from the control run without assimilation (FR). The problems of performance evaluations of operational ocean models were
addressed by Hernandez et al. (2015). In our study, the option of withholding the observations was performed: it was evaluated
how much the DA result will change if DA is performed using 50% of the available data (Gregg et al., 2009). The present
implementation of EOF DA used about 13 k observational averages over coarse 5' N × 10' E grid. The reconstruction procedure
425   by Eqs. (1)–(2) has no direct connection to the ongoing modelling (although it includes statistical results from longer model
runs) and the fields of ψ^o in Eqs. (3)–(4) are the only link where observations enter the DA process. The experiments which
took every second available observation "box" into account (this resulted in mean sampling interval along ship tracks about
20 km instead of 10 km) revealed that performing DA during the study period with reduced data set (6.5 k averaged observation
data instead of 13 k) changed RMSD of SST by only 1% and of SSS by 2%. It was evaluated over the full time span and
430   domain using 182 k coarse grid cells; it was not possible to distinguish correlation between the data sets to be different from
unity. Consequently, for our large-scale approach DA results are robust to the reasonable variation of data amount and we used
FB data for reference in the skill evaluations.

Evaluated skill metrics are presented in Table 2. Only those fine grid points were used for metrics calculation, which had
435   respective value of FerryBox observations on the same day. Wet-points of the model without corresponding observation value
were left out from the procedure.

The skill properties presented in Table 2 reflect that DA improves the model performance significantly: RMSD of SST was
reduced by 22% and SSS by 34%, compared to the control run. From DA01 to DA02, slight improvement of DA performance
440   was observed, therefore we adopted DA02 as the major result. Spatial pattern of RMSD change between the DA and FR (Fig.
8) reveals that larger reduction rates (up to 50%), both for SST and SSS, were found in the observation-covered areas in the
Gulf of Finland. Too cold waters produced by FR near the northern coast of the Gulf of Finland were effectively corrected by
DA (see also Fig. 4), therefore highest improvement percentage scores were detected in this region. Near the western open
boundary, non-assimilated SST and SSS values of the larger model were advected into the area, therefore RMSD reduction
445   was small, or even negative for SSS.

The applied EOF DA method does not assimilate mesoscale variability. Applying the weekly average statistics like Zujev and
Elken (2018), further reduced RMSD by 13% for SST and 9% for SSS, compared to the daily data in Table 2. Weekly statistics

suppresses the mesoscale variability and reveals better match between the DA and the observations. DA decreased the bias,
450 especially for SSS. At the same time, correlation of SSS between DA and observations increased considerably. We may
conclude that DA made major improvement in modelling of SSS. Still, RMSD to the observations makes 62% of observed
standard deviations, calling for further improvements in modelling SSS in the Baltic Sea. Modelling of SST is more accurate
than SSS already without DA: RMSD of control run (FR) makes 18% of the standard deviation of observations for SST and
94% for SSS.

4 Discussion

Baltic Sea is considered as one of the most studied marine areas in the world (e.g. Andersen et al., 2017). However, the large
observational data sets are distributed unevenly. If we divide our study area into 744 eddy-averaging 5' N × 10' E grid cells,
then during the study period 330 k FerryBox observations covered only 18% of the sea region. Shipborne monitoring added
more 8% coverage of the area, but with much smaller frequency of sampling. Having in mind that the ocean models tend to
460 deviate in the NE Baltic from the observations not only by constant bias but also for large-scale and longer-term response,
introduction of non-local, region wide data assimilation is of high importance.

It is interesting to consider how our statistical evaluations of model and DA performance, given in Table 2, compare with other
Baltic Sea studies. For remote sensing versus in situ reference, Kozlov et al. (2014) have found RMSD 1.31 °C in the Curonian
465 Lagoon. Uiboupin and Laanemets (2015) have estimated RMSD of various satellite products to FerryBox in the Gulf of Finland
from 0.29 to 0.98 °C. Our control run gave RMSD 0.72 °C. Golbeck et al. (2015) compared SST from 13 models with satellite
data and found in the Baltic Sea yearly RMSD for SST 0.65–0.87 °C. They found larger relative spread of SSS ensemble
members than of SST: deviations in the Gulf of Finland between the models were nearly up to 1 g kg^{-1}, while the average SSS
is only about 4 g kg^{-1}. Unfortunately, there were not enough validating observations for SSS available. Fu et al. (2011) found
470 for the control run RMSD for SST even larger, 1.0 °C, based on satellite observations. They also used DA with ensemble
optimal interpolation and found that DA reduced RMSD between the forecasts and observations by 25% for SST and 34% for
SSS. With our simpler and less computationally demanding EOF DA technique, similar RMSD reductions have been obtained
(Sect. 3.2.4) compared to earlier studies.

475 We have developed and tested an EOF-based relaxation technique where the large-scale observed fields to be assimilated are
pre-calculated independently from the ongoing model. From sparse observations, it is possible to estimate the amplitudes of
only the gravest, large-scale EOF modes. The EOF DA method handles large-scale features over the sea basin(s), like change
of mean SST, SSS and their gradients, including differential heating in coastal and offshore areas, major patterns from
upwelling, and spreading of river discharge. The method can work well with irregular data, but cannot resolve mesoscale
480 features in the areas of dense observations, because the EOF amplitudes of higher modes get noisy, according to our

experiments. Optimal interpolation, successive corrections and similar methods assume usually localized covariance and/or radius of influence (e.g. Axell and Liu, 2016); they work well in resolving mesoscale in dense sampling areas, but regions of rare observations remain unaffected by DA. For mesoscale range, in our study area there are only satellite observations of surface variables available. They were omitted from our study, since salinity as a variable of primary interest can be presently

485   determined in the Baltic only in situ. It is possible to implement on top of EOF DA more traditional localized DA methods to assimilate mesoscale data when and where such data are available. Studies on using EOF DA for handling large-scale data are also ongoing in the UK Met Office by Daniel Lea (Haines, 2018).

The EOF DA method has some practical advantages. Firstly, it has small computational effort compared to the localized

490   methods like optimal interpolation etc. Secondly, intermediate results are in the form of maps that are easily understandable and can be checked visually or taught to be analyzed by artificial intelligence. For optimizing the observational data needs, the concept of OSE (Observing System Experiments) that checks various data configurations for DA performance, are highly on the agenda. Since the quality of DA and forecast are primarily determined by the quality of EOF reconstruction (when extensive mesoscale observations are not available), then it would be possible to save a significant amount of computing power and

495   perform most of the experiments using orthogonality of the EOF basis vectors.

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 much variance is explained by the major EOF modes. There are a number of questions that may be addressed, like: What is

500   the minimal amount of observations needed to produce decent results? What areas are reconstructed with higher accuracy with given observation design, nearshore, offshore, open basins? What areas are most problematic to reconstruct, complicated coastline, straits and channels, semi-enclosed basins, regions of river influence? Are there some specific locations that can be used as a proxy for larger regions? Is it possible to measure SST/SSS just at these points in order to give enough input for successful reconstruction?

505   **5 Conclusions**

The present study was aimed to implement EOF-based statistical reconstruction technique into the data assimilation of the forecast model, and to study the feasibility of such assimilation method. Gridded EOF modes were determined from the 5-year long model results. "Observational" EOF amplitudes were found each day to minimize the RMSD between the reconstructed and observed values at the observation points, using time-dependent technique where both the amplitudes and their time rate

510   of change were searched for the best fit. In this procedure, a time window of 30 days was selected that ensured acceptable SST and SSS reconstruction patterns by three gravest EOF modes throughout the whole study period from 1 May to 31 December 2015. The study used about 330 k (thousand) FerryBox observations along four ship tracks from 1 May to 31 December 2015,

and 370 observations from research vessels. Statistically gridded observations were daily assimilated into the model by the relaxation techniques, using restoring times of 5 and 10 days.

515

The tested EOF-based data assimilation (DA) method decreased RMSD of surface temperature (SST) and salinity (SSS) in the NE Baltic Sea by 22% and 34%, respectively, compared to the control run without DA. Using the observation-estimated amplitudes of the pre-calculated gravest model-based EOF modes, the method is able to follow on the regular grid the pointwise observed temporal changes of the mean state and of the major basin-scale gradients. DA with EOF reconstruction technique

520 was found feasible for further implementation studies, since: 1) the method that works on the large-scale patterns (mesoscale features are neglected by taking only the gravest EOF modes) improves the high-resolution model performance by comparable or even better degree than in the other published studies, 2) the method is computationally effective.

Code and data availability

The model code has been developed by the Baltic MFC partners. Presently it is frozen and not anymore developed. The DA

525 scripts and demonstrated model results can be requested by contacting the corresponding author. All the used observational data are freely available as described in Sect. 2.2.

Author contribution

MZ carried out DA experiments and performed analysis of the results. JE worked on theoretical aspects and performed gridded reconstruction of observations. PL worked with the circulation model. All authors contributed in discussion, planning and

530 writing.

Acknowledgements

The study was supported by the PhD program for Mihhail Zujev and the institutional research funding. A larger BAL MFC team did development of the HBM model within the EU projects MyOcean, MyOcean2 and MyOcean-FO. There is an ongoing activity to develop and maintain Baltic monitoring and forecasting services within the CMEMS. This cooperation is highly

535 acknowledged.

References

Alenius, P., Myrberg, K. and Nekrasov, A.: The physical oceanography of the Gulf of Finland: a review. Boreal Environ. Res, 3(2), 97–125, 1998.

Andersen, J.H., Carstensen, J., Conley, D.J., Dromph, K., Fleming-Lehtinen, V., Gustafsson, B.G., Josefson, A.B., Norkko,

540   A., Villnäs, A. and Murray, C.: Long-term temporal and spatial trends in eutrophication status of the Baltic Sea. Biological Reviews, 92(1), 135–149, doi: 10.1111/brv.12221, 2017.

Axell, L. and Liu, Y.: Application of 3-D ensemble variational data assimilation to a Baltic Sea reanalysis 1989–2013. Tellus A: Dynamic Meteorology and Oceanography, 68(1), 24220, http://dx.doi.org/10.3402/tellusa.v68.24220, 2016.

Berg, P., and Poulsen, J. W.: Implementation details for HBM. DMI Technical report No. 12-11. Copenhagen, 2012.

545   Buizza, R., Brönnimann, S., Haimberger, L., Laloyaux, P., Martin, M.J., Fuentes, M., Alonso-Balmaseda, M., Becker, A., Blaschek, M., Dahlgren, P. and De Boisseson, E.: The EU-FP7 ERA-CLIM2 project contribution to advancing science and production of earth system climate reanalyses. Bulletin of the American Meteorological Society, 99(5), 1003–1014, doi: 10.1175/BAMS-D-17-0199.1, 2018.

Carrassi, A., Bocquet, M., Bertino, L. and Evensen, G.: Data assimilation in the geosciences: An overview of methods, issues,

550   and perspectives. Wiley Interdisciplinary Reviews: Climate Change, 9(5), e535, doi: 10.1002/wcc.535, 2018.

Elken, J., Zujev, M. and Lagemaa, P.: Reconstructing sea surface temperature and salinity fields in the northeastern Baltic from observational data, based on sub-regional Empirical Orthogonal Function (EOF) patterns from models. In 2018 IEEE/OES Baltic International Symposium (BALTIC) (pp. 1–8). IEEE, doi: 10.1109/BALTIC.2018.8634845, 2018.

Elken, J., Zujev, M., She, J. and Lagemaa, P.: Reconstruction of large-scale sea surface temperature and salinity fields using

555   sub-regional EOF patterns from models. Frontiers in Earth Science, 7, doi: 10.3389/feart.2019.00232, 2019.

Fu, W., She, J. and Zhuang, S.: Application of an Ensemble Optimal Interpolation in a North/Baltic Sea model: Assimilating temperature and salinity profiles. Ocean Modelling, 40(3–4), 227–245, doi:10.1016/j.ocemod.2011.09.004, 2011.

Fujii, Y., Remy, E., Zuo, H., Oke, P.R., Halliwell, G.R., Gasparin, F., Benkiran, M., Loose, N., Cummings, J., Xie, J. and Xue, Y.: Observing system evaluation based on ocean data assimilation and prediction systems: on-going challenges and future

560   vision for designing/supporting ocean observational networks. Frontiers in Marine Science, 6, doi: 10.3389/fmars.2019.00417, 2019.

Golbeck, I., Li, X., Janssen, F., Brüning, T., Nielsen, J.W., Huess, V., Söderkvist, J., Büchmann, B., Siiriä, S.M., Vähä-Piikkiö, O., Hackett, B., Kristensen, N., Engedahl, H., Blockey, E., Sellar, A., Lagemaa, P., Ozer, J., Legrand, S., Ljungemyr, P. and Axell, L.: Uncertainty estimation for operational ocean forecast products—a multi-model ensemble for the North Sea and the

565   Baltic Sea. Ocean Dynamics, 65(12), 1603–1631, doi: 10.1007/s10236-015-0897-8, 2015.

Goodliff, M., Bruening, T., Schwichtenberg, F., Li, X., Lindenthal, A., Lorkowski, I. and Nerger, L.: Temperature assimilation into a coastal ocean-biogeochemical model: assessment of weakly and strongly coupled data assimilation. Ocean Dynamics, 69(10), 1217–1237, doi: 10.1007/s10236-019-01299-7, 2019.

Gregg, W.W., Friedrichs, M.A., Robinson, A.R., Rose, K.A., Schlitzer, R., Thompson, K.R. and Doney, S.C.: Skill assessment

570   in ocean biological data assimilation. Journal of Marine Systems, 76(1-2), 16–33, doi: 10.1016/j.jmarsys.2008.05.006, 2009.

Haines, K.: Ocean Reanalyses. New Frontiers in Operational Oceanography, 545–562, doi :10.17125/gov2018.ch19, 2018.

Hernandez, F., Blockley, E., Brassington, G.B., Davidson, F., Divakaran, P., Drévillon, M., Ishizaki, S., Garcia-Sotillo, M., Hogan, P.J., Lagemaa, P. and Levier, B.: Recent progress in performance evaluations and near real-time assessment of operational ocean products. Journal of Operational Oceanography, 8(sup2), s221–s238, doi:

575   10.1080/1755876X.2015.1050282, 2015.

Holland, W.R. and Malanotte-Rizzoli, P.: Assimilation of altimeter data into an ocean circulation model: Space versus time resolution studies. Journal of Physical Oceanography, 19(10), 1507–1534, 1989.

Huess, V.: Product User Manual For Baltic Sea Physical Analysis and Forecasting Product, http://marine.copernicus.eu/documents/PUM/CMEMS-BAL-PUM-003-006.pdf, assessed 2 May 2020, 2020.

580   Jevrejeva, S., Drabkin, V.V., Kostjukov, J., Lebedev, A.A., Leppäranta, M., Mironov, Y.U., Schmelzer, N. and Sztobryn, M.: Baltic Sea ice seasons in the twentieth century. Climate Research, 25(3), 217–227, 2004.

Johansson, J.: Total and regional runoff to the Baltic Sea. Baltic Sea environment fact sheet, available at: http://www. helcom. fi/baltic-sea-trends/environment-fact-sheets/, last access: April 2020, 2017.

Kõuts, T. and Omstedt, A.: Deep water exchange in the Baltic Proper. Tellus A: Dynamic Meteorology and Oceanography,

585   45(4), 311–324, 1993.

Kozlov, I., Dailidienė, I., Korosov, A., Klemas, V. and Mingėlaitė, T.: MODIS-based sea surface temperature of the Baltic Sea Curonian Lagoon. Journal of Marine Systems, 129, 157–165, doi:10.1016/j.jmarsys.2012.05.011, 2014.

Laanemets, J., Zhurbas, V., Elken, J. and Vahtera, E.: Dependence of upwelling-mediated nutrient transport on wind forcing, bottom topography and stratification in the Gulf of Finland: model experiments. Boreal environment research, 14(1), 213–225,

590   2009.

Lagemaa, P.: Operational forecasting in Estonian marine waters (Vol. 128). TUT Press, 2012.

Le Traon, P.Y., Reppucci, A., Alvarez Fanjul, E., Aouf, L., Behrens, A., Belmonte, M., Bentamy, A., Bertino, L., Brando, V.E., Kreiner, M. and Benkiran, M.: From observation to information and users: the Copernicus Marine Service perspective. Frontiers in Marine Science, 6, doi: 10.3389/fmars.2019.00234, 2019.

595   Liblik, T., Laanemets, J., Raudsepp, U., Elken, J. and Suhhova, I.: Estuarine circulation reversals and related rapid changes in winter near-bottom oxygen conditions in the Gulf of Finland, Baltic Sea. Ocean Science, 9(5), 917–930, doi: 10.5194/os-9-917-2013, 2013.

Lilover, M.J., Lips, U., Laanearu, J. and Liljebladh, B.: Flow regime in the Irbe Strait. Aquatic Sciences, 60(3), 253–265, 1998.

Liu, Y. and Fu, W.: Assimilating high-resolution sea surface temperature data improves the ocean forecast potential in the

600   Baltic Sea. Ocean Science, 14(3), 525–541, doi: 10.5194/os-14-525-2018, 2018.

Liu, Y., Meier, M. and Eilola, K.: Nutrient transports in the Baltic Sea-results from a 30-year physical-biogeochemical reanalysis. Biogeosciences, 14(8), 2113–2131, doi: 10.5194/bg-14-2113-2017, 2017.

Martin, M.J., Balmaseda, M., Bertino, L., Brasseur, P., Brassington, G., Cummings, J., Fujii, Y., Lea, D.J., Lellouche, J.M., Mogensen, K. and Oke, P.R.: Status and future of data assimilation in operational oceanography. Journal of Operational

605   Oceanography, 8(sup1), s28–s48, http://dx.doi.org/10.1080/1755876X.2015.1022055, 2015.

Meier, H.E.M., Eilola, K., Almroth-Rosell, E., Schimanke, S., Kniebusch, M., Höglund, A., Pemberton, P., Liu, Y., Väli, G. and Saraiva, S.: Disentangling the impact of nutrient load and climate changes on Baltic Sea hypoxia and eutrophication since 1850. Climate dynamics, 53(1–2), 1145–1166, doi: 10.1007/s00382-018-4296-y, 2019.

Moore, A.M. and Reason, C.J.: The response of a global ocean general circulation model to climatological surface boundary conditions for temperature and salinity. Journal of Physical Oceanography, 23(2), 300–328, 1993.

Moore, A.M., Martin, M.J., Akella, S., Arango, H., Balmaseda, M.A., Bertino, L., Ciavatta, S., Cornuelle, B., Cummings, J., Frolov, S. and Lermusiaux, P.: Synthesis of ocean observations using data assimilation for operational, real-time and reanalysis systems: A more complete picture of the state of the ocean. Frontiers in Marine Science, 6, doi: 10.3389/fmars.2019.00090, 2019.

Omstedt, A. and Axell, L.B.: Modeling the variations of salinity and temperature in the large Gulfs of the Baltic Sea. Continental Shelf Research, 23(3–4), 265–294, doi: 10.1016/S0278-4343(02)00207-8, 2003.

Placke, M., Meier, H.E., Gräwe, U., Neumann, T., Frauen, C. and Liu, Y.: Long-term mean circulation of the Baltic Sea as represented by various ocean circulation models. Frontiers in Marine Science, 5, doi: 10.3389/fmars.2018.00287, 2018.

Raudsepp, U. and Elken, J.: Application of the Bryan-Cox-Type Ocean Model to reproduce synoptic and mesoscale variability of the Irbe Strait salinity front. Deutsche Hydrografische Zeitschrift, 51(4), 477–488, 1999.

Ravichandran, M., Behringer, D., Sivareddy, S., Girishkumar, M.S., Chacko, N. and Harikumar, R.: Evaluation of the global ocean data assimilation system at INCOIS: the tropical Indian Ocean. Ocean Modelling, 69, 123–135, doi: 10.1016/j.ocemod.2013.05.003, 2013.

Rodhe, J.: The Baltic and North Seas: a process-oriented review of the physical oceanography. The sea, 11, 699–732, 1998.

Savchuk, O.P.: Large-scale nutrient dynamics in the Baltic Sea, 1970–2016. Frontiers in Marine Science, 5, doi: 10.3389/fmars.2018.00095, 2018.

She, J.: Assessment of Baltic Sea observations for operational oceanography. In Proceedings of the 8th EuroGOOS International Conference, eds E. Buch, V. Fernández, D. Eparkhina, P. Gorringe, and G. Nolan (Bergen: EuroGOOS) (pp. 79–87), 2018.

She, J., Meier, M., Darecki, M., Gorringe, P., Huess, V., Kouts, T., Reissmann, J.H. and Tuomi, L.: Baltic Sea Operational Oceanography-A Stimulant for Regional Earth System Research. Frontiers in Earth Science, 8, doi: 10.3389/feart.2020.00007, 2020.

Stow, C.A., Jolliff, J., McGillicuddy Jr, D.J., Doney, S.C., Allen, J.I., Friedrichs, M.A., Rose, K.A. and Wallhead, P.: Skill assessment for coupled biological/physical models of marine systems. Journal of Marine Systems, 76(1–2), 4–15, doi: 10.1016/j.jmarsys.2008.03.011, 2009.

Stramska, M. and Białogrodzka, J.: Spatial and temporal variability of sea surface temperature in the Baltic Sea based on 32-years (1982–2013) of satellite data. Oceanologia, 57(3), 223–235, doi:10.1016/j.oceano.2015.04.004, 2015.

Tuomi, L., She, J., Lorkowski, I., Axell, L., Lagemaa, P., Schwichtenberg, F., Huess, V.: Overview of CMEMS BAL MFC Service and Developments. Proceedings of the Eight EuroGOOS International Conference, pp. 261–267, ISBN 978-2-

640   9601883-3-2, 2017.

Uiboupin, R. and Laanemets, J.: Upwelling characteristics derived from satellite sea surface temperature data in the Gulf of Finland, Baltic Sea. Boreal Environment Research, 14(2), 297–304, 2009.

Uiboupin, R. and Laanemets, J.: Upwelling parameters from bias-corrected composite satellite SST maps in the Gulf of Finland (Baltic Sea). IEEE Geoscience and Remote Sensing Letters, 12(3), 592–596, doi: 10.1109/LGRS.2014.2352397, 2015.

645   Yurkovskis, A., Wulff, F., Rahm, L., Andruzaitis, A. and Rodriguez-Medina, M.: A nutrient budget of the Gulf of Riga; Baltic Sea. Estuarine, Coastal and Shelf Science, 37(2), 113–127, doi: 10.1006/ecss.1993.1046, 1993.

Zujev, M. and Elken, J.: Testing marine data assimilation in the northeastern Baltic using satellite SST products from the Copernicus Marine Environment Monitoring Service. Proceedings of the Estonian Academy of Sciences, 67(3), 217–230, https://doi.org/10.3176/proc.2018.3.03, 2018.

650

Table 1: FerryBox data from 1 May to 31 December 2015 in the NE Baltic used in the present study.

Ship	Main route	Operating institute	Number of initial observations
Baltic Queen	Tallinn – Helsinki	Marine Systems Institute, Tallinn University of Technology	63 368
FinnMaid	Helsinki (Vuosaari) – Travemünde	Finnish Environment Institute	142 235
Silja Serenade	Helsinki – Mariehamn – Stockholm	Finnish Environment Institute	60 228
Victoria	Tallinn – Mariehamn – Stockholm	Estonian Marine Institute, University of Tartu	65 037

655

Table 2: Statistics of daily data in 0.5' N × 1' E grid cells with FerryBox (FB) observations: free model run without data assimilation (FR), data assimilation DA01 (observation weight 0.1), DA02 (weight 0.2) and FB. Bias, RMSD and correlation are taken with reference to FB.

	FR	DA01	DA02	FB
SST (°C)				
Mean	12.03	12.15	12.25	12.48
Standard deviation	3.98	3.92	3.93	3.97
Bias	-0.45	-0.33	-0.23	0
RMSD	0.72	0.59	0.56	0
Correlation	0.98	0.99	0.99	1.00
SSS (g kg^{-1})				
Mean	5.61	5.79	5.85	5.93
Standard deviation	0.35	0.29	0.31	0.37
Bias	-0.31	-0.14	-0.08	0
RMSD	0.35	0.24	0.23	0
Correlation	0.52	0.76	0.78	1.00

660

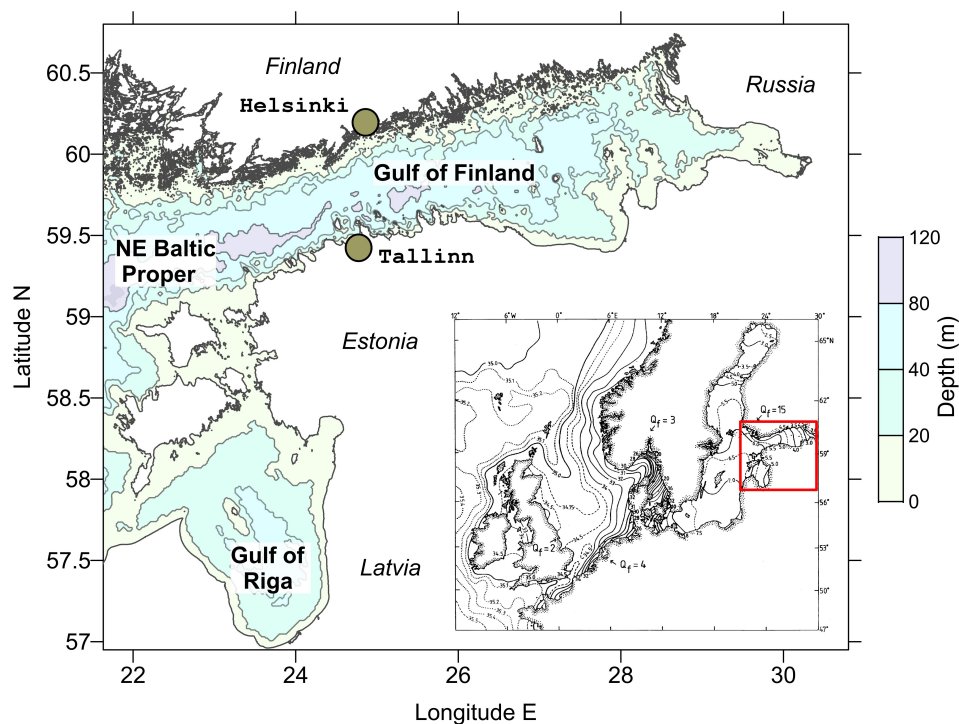



Figure 1: Map of the study area in the NE Baltic with depth contours. Shown are the sea areas of Gulf of Finland, Gulf of Riga and part of the NE Baltic Proper. Insert presents the map of surface salinity of the Baltic and North seas by Rohde (1998). The arrows present the mean basin-wide river discharges in 1000 m³ s⁻¹. Location of our study area is given on the insert by a red box.

665

670

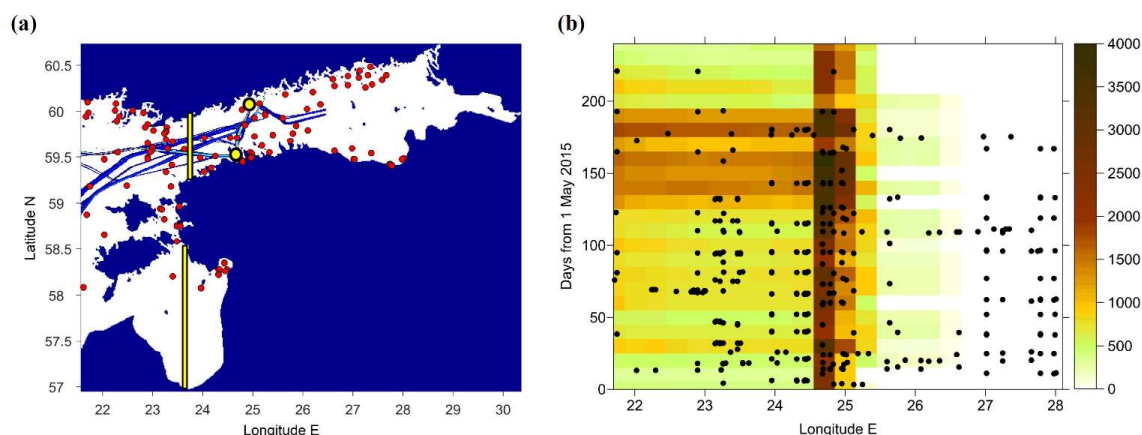
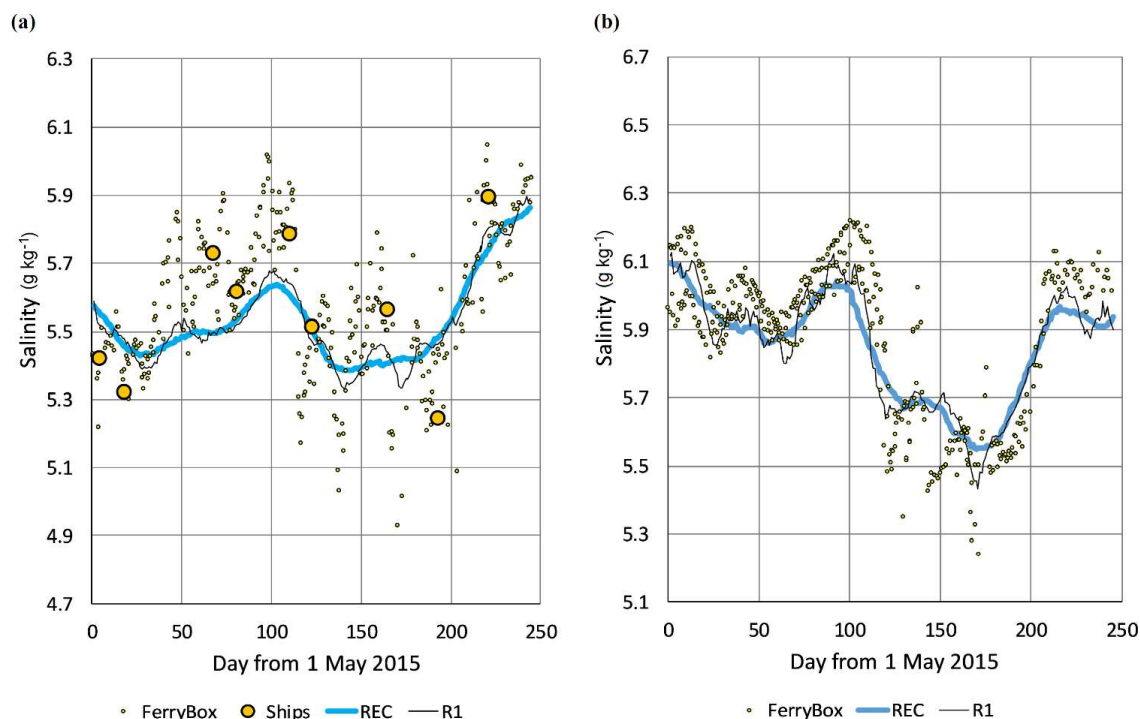



Figure 2: Distribution of observations. (a) Map of FerryBox observation points along ship tracks (blue) and shipborne monitoring observations (red) over the study period. Shown are also the locations for time-latitude graphs and time series (black contours with yellow background. (b) Observation frequency over longitude and time. FerryBox data are shown by colour image; each image cell presents the number of initial observations over intervals of 10 days and 18' E longitude. Shipborne observations are shown by black
675   **dots.**

680 **Figure 3: Salinity time series at locations (a) 59.8383° N, 24.8383° E (HELCOM station F3) and (b) 59.794° N, 24.822° E, during the study period. Shown by dots are the observations from FerryBox and from ships (a, monitoring). Reconstructed time series, made by the time-dependent method, are given by solid lines: REC – basic option with 30 days interval, all observations in window were kept as they are; R1 – the same as previous but with time interval 10 days.**

685

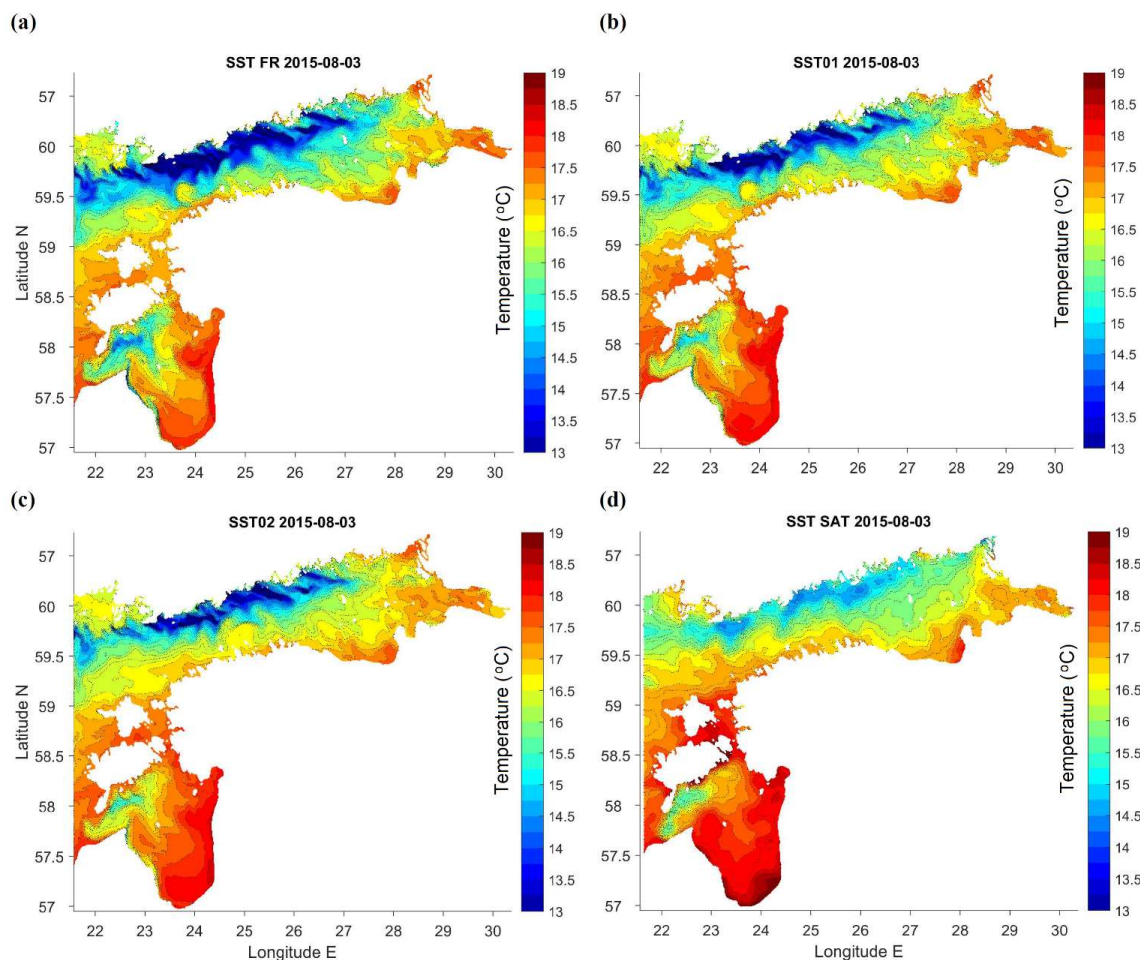
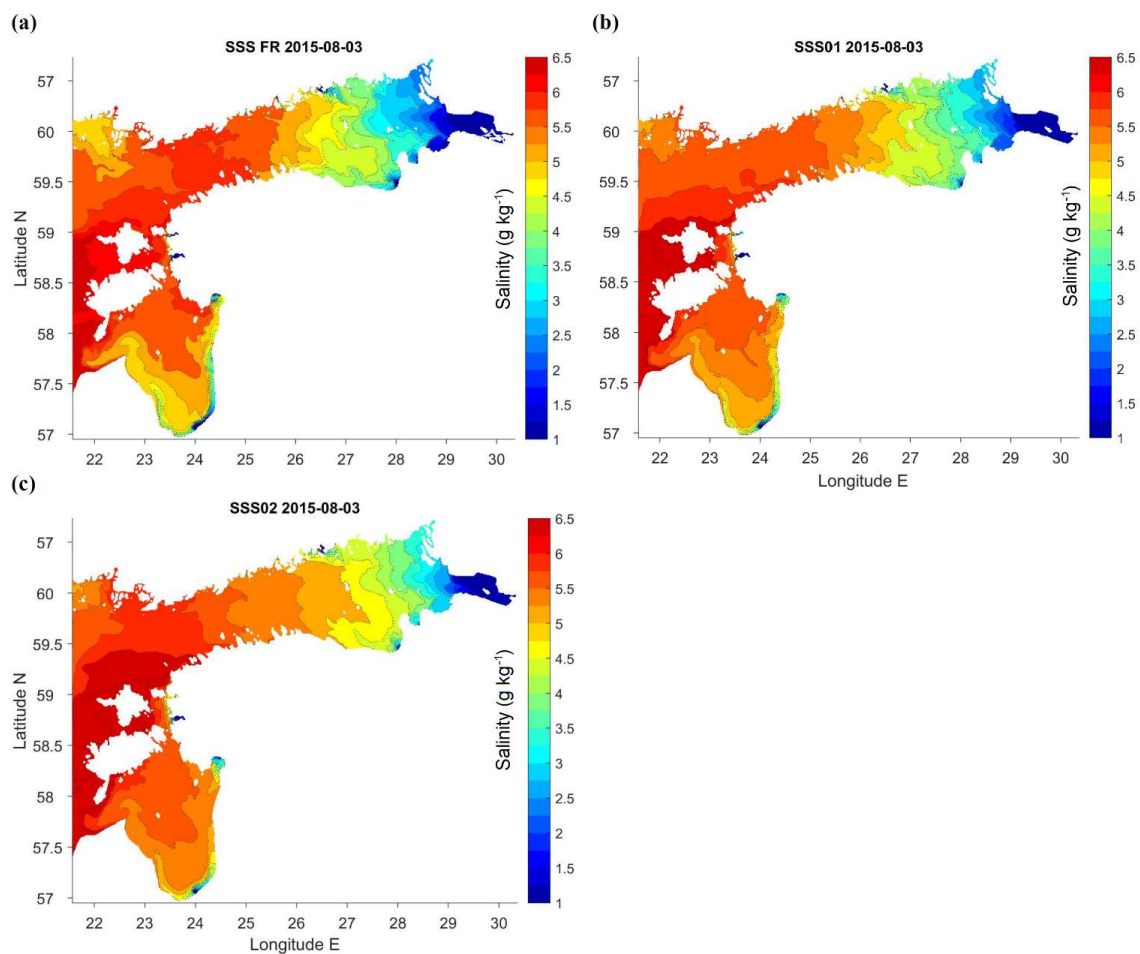



Figure 4: Maps (longitude E, latitude N) of SST in the study area on 3 August 2015: (a) free model run without DA, (b) DA with
690 relaxation time 10 days (weight of observations 0.1), (c) DA with relaxation time 5 days (weight 0.2), (d) satellite observations.

695 **Figure 5: Maps (longitude E, latitude N) of SSS in the study area on 3 August 2015: (a) free model run without DA, (b) DA with relaxation time 10 days (weight of observations 0.1), (c) DA with relaxation time 5 days (weight 0.2).**

700

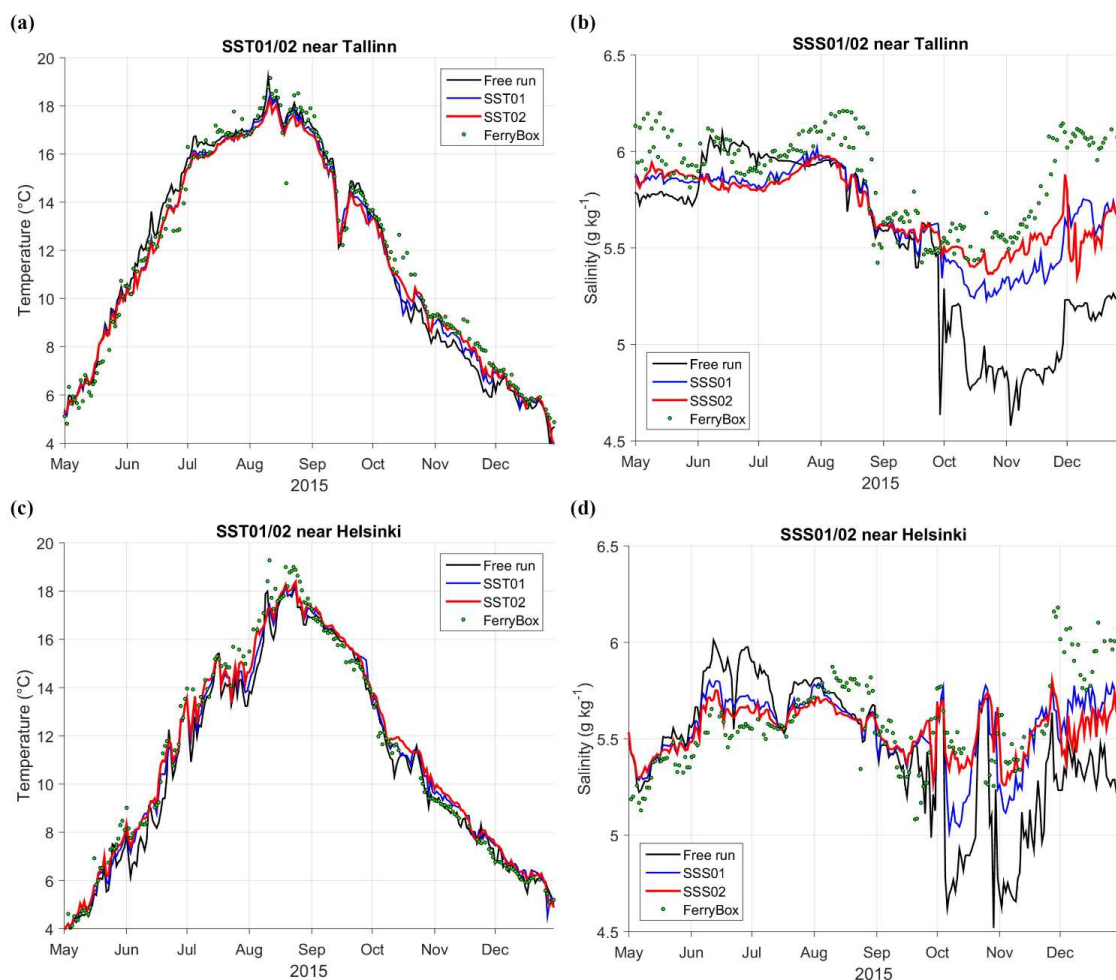
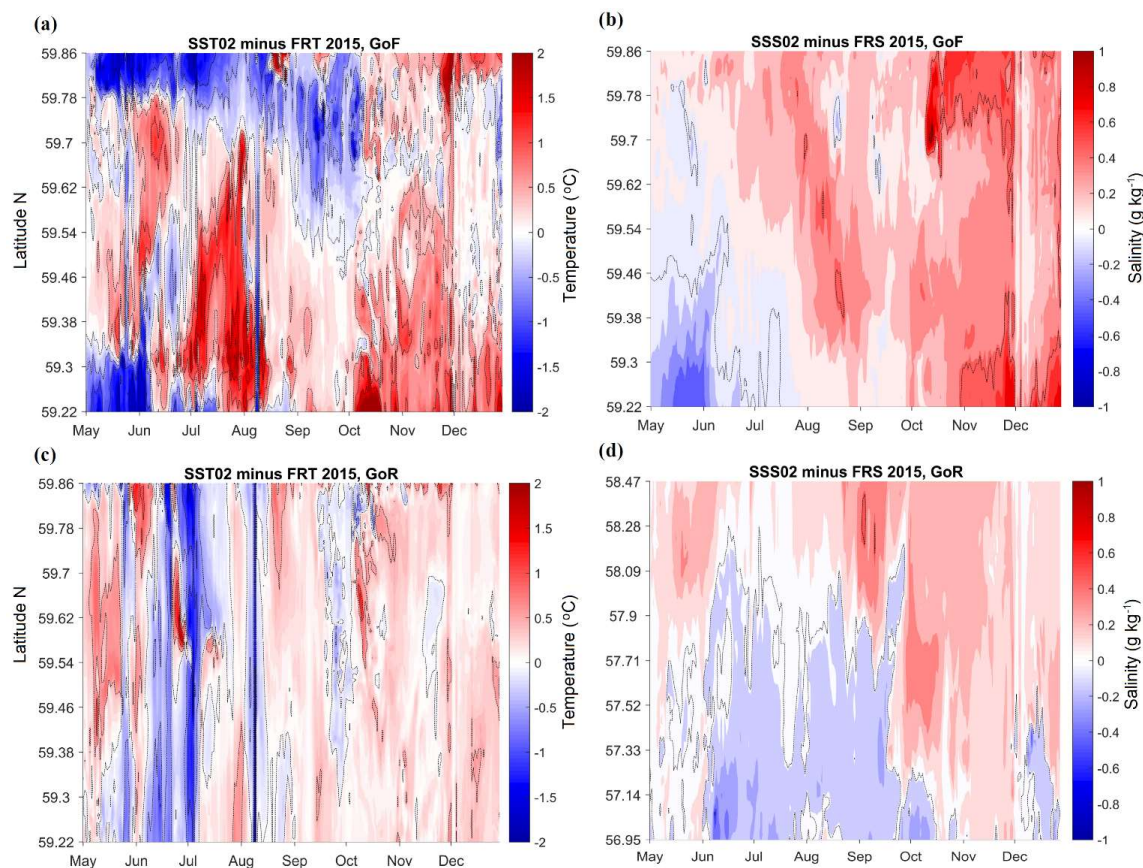



Figure 6: Time series of SST (a, c) and SSS (b, d) near Tallinn (a, b, 59.4833° N, 24.7667° E) and Helsinki (c, d, 59.9500° N, 24.8833° E), locations shown in Fig. 2a. FerryBox data are shown by dots, black lines represent control run without DA, red lines correspond to DA with relaxation time 5 days (weight of observations 0.2), blue lines for 10 days (weight 0.1).

705

710   **Figure 7: Time (months of 2015) versus latitude (N) contour graph of DA anomalies of SST (a, c) and SSS (b, d) in reference to the control run (FR) without data assimilation; at longitudes 23.7166° E (a, b, Gulf of Finland) and 23.5333° E (c, d, Gulf of Riga), locations shown in Fig. 2a. DA data are given for relaxation time 5 days (weight of observations 0.2).**

715

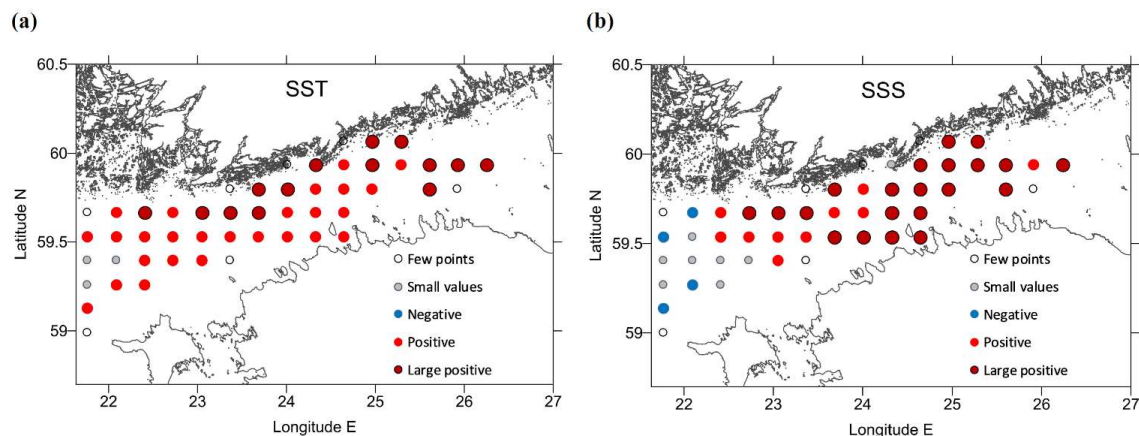



Figure 8: Improvement of RMSD of DA compared to that of FR, both taken in reference to 110 k FerryBox observations. Comparison is made for 20 x 20 grid cells (10' N × 20' E) for SST (a) and SSS (b) over the whole study period. Legend codes: few points - less than 100 observations in a box, small values - absolute percentage change less than 10%, negative - DA RMSD growth more than 10%, positive - DA improvement (RMSD reduction) from 10% to 30%, large positive - improvement more than 30%.

720