Using model-based sub-regional EOF patterns to reconstruct temperature and salinity fields from observations

Jüri Elken, Mihhail Zujev, Priidik Lagemaa
Department of Marine Systems, Tallinn University of Technology

Contents:

Background

How far can we extend the data from modern FerryBoxes and automatic offshore buoy stations?
In combination with other observations?

The study came out from **data assimilation experiments**:

Covariances of SST and SSS show **large values over long distances**
= large-scale patterns dominate variability in the sub-region

$\Longrightarrow$ this creates technical problems in data assimilation using optimal interpolation, 3D-VAR etc

Subtracting seasonal harmonics and/or climatology **does not remove**
large covariance scales, large-scale patterns still present

$\Rightarrow$ optimal interpolation usually assumes fast decay of covariance with space lag

Analysis of covariances calculated from the model data (average over time) suggest
EOF analysis

??? Leading EOF modes can be evaluated by calculating EOF amplitudes by limited amount of observational data = **interpolation of observations**

Methods (1): Data and covariances

Take a set of gridded (model) data covering quite long time span: assume that model is „good“

M spatial points, N moments in time

$\mathbf{X}$ **space – time data** $M \times N$

spatial data are aligned linearly, each element has geographical coordinates

Average over time

space-dependent temporal mean $\overline{\mathbf{x}}_m$

$\mathbf{X}'$ **space – time data deviations** $M \times N$ from $\overline{\mathbf{x}}_m$

$\mathbf{x}'_i$ **spatial data vector** at time slice i (*usually transformation of gridded data at specific time*)

Covariance: average over time of all pairs of spatial data point products

Covariance matrix $\mathbf{B} = \frac{1}{N-1}\mathbf{X}'^{\mathrm{T}}\mathbf{X}'$

Methods (2): Empirical Orthogonal Functions (EOF)

Covariance $\mathbf{B} = \frac{1}{N-1}\mathbf{X}'^{\mathrm{T}}\mathbf{X}'$

Eigenvalue problem $|\mathbf{B} - \lambda\mathbf{I}| = 0$

$\mathbf{E}$ **eigenvectors** (empirical orthogonal functions) are found from $\mathbf{BE} = \mathbf{\Lambda E}$

$\mathbf{\Lambda}$ is diagonal matrix containing eigenvalues λ_k

columns of $M \times M$ matrix $\mathbf{E}$ are space-dependent eigenvectors $\mathbf{e}_k$ of mode k

$\mathbf{A}$ **amplitude** matrix, $M \times N$ columns are time-dependent amplitudes $\mathbf{a}_i$

Decomposition of data by time-dependent amplitudes and space-dependent eigenvectors

$\boxed{\tilde{\mathbf{a}}_i = \mathbf{\Lambda a}_i = \mathbf{E}^{\mathrm{T}}\mathbf{x}'_i}$ $\boxed{\mathbf{x}'_i = \mathbf{E}\tilde{\mathbf{a}}_i}$ alternatives $\mathbf{x}'_i = \mathbf{E\Lambda a}_i$ $\mathbf{X}' = \mathbf{E\Lambda A}$

orthonormality $\mathbf{E}^{\mathrm{T}}\mathbf{E} = \mathbf{I}$ $\mathbf{e}_i\mathbf{e}_j = \delta_{i,j}$ $\mathbf{a}_i\mathbf{a}_j = \delta_{i,j}$ $\tilde{\mathbf{a}}_i\tilde{\mathbf{a}}_j = \lambda_i^2\delta_{i,j}$

Methods (3): Interpolation with EOF

Observations $\mathbf{y}$ are on a different set of K points than $\mathbf{x}_i'$ $K < M$

$\mathbf{H}$ transforms gridded data $\mathbf{x}_i'$ to the observation points, $\mathbf{H}\mathbf{x}_i'$

$\mathbf{H}\hat{\mathbf{x}}' = \mathbf{H}\mathbf{E}\,\hat{\mathbf{a}}$ $\hat{\mathbf{a}}$ is „observational“ amplitude calculated using „full“ $\mathbf{e}_k$

Least-squares error minimization $\left\|\mathbf{y} - \mathbf{H}\mathbf{x}_i'\right\|^2 = \left\|\mathbf{y} - \mathbf{H}\mathbf{E}\hat{\mathbf{a}}_i\right\|^2$

gives system of equations $\mathbf{H}^{\mathrm{T}}\mathbf{E}^{\mathrm{T}}\mathbf{H}\mathbf{E}\hat{\mathbf{a}}_i = \mathbf{H}^{\mathrm{T}}\mathbf{E}^{\mathrm{T}}\mathbf{y}$

Solution for „observational“ amplitudes is $\hat{\mathbf{a}}_i = \left(\mathbf{H}^{\mathrm{T}}\mathbf{E}^{\mathrm{T}}\mathbf{H}\mathbf{E}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{E}^{\mathrm{T}}\mathbf{y}$

Gridded interpolation is obtained by $\hat{\mathbf{x}}_i' = \mathbf{E}\hat{\mathbf{a}}_i$

We take only first L eigenmodes

Methods (4): Interpolation with EOF (continued)

Alternative form of equations to find $\hat{\mathbf{a}}$:

$$\sum_{k=1}^{K}\left(y_k - \sum_{i=1}^{L}\hat{a}_i\hat{e}_i^k\right)^2 \Rightarrow \min$$

Here $\hat{e}_i^k$ is the value of i -th eigenvector in the observation point k and y_k are observed values.

Solve the system of L linear equations

$$\mathbf{D}\hat{\mathbf{a}} = \mathbf{h}$$

where

$$D_{ij} = \sum_{k=1}^{K}\hat{e}_i^k\,\hat{e}_j^k \qquad h_j = \sum_{k=1}^{K} y_k\,\hat{e}_j^k$$

Original EOF amplitudes are determined over full set of grid points. „Observational“ amplitudes are determined over much smaller number of space points and may be rather uncertain.

CAUTION: with bad configuration of observation points, „observational“ EOF **amplitudes of particular modes may get larger than limits** determined from „full“ statistics. **These and higher modes need to be omitted from interpolation.**

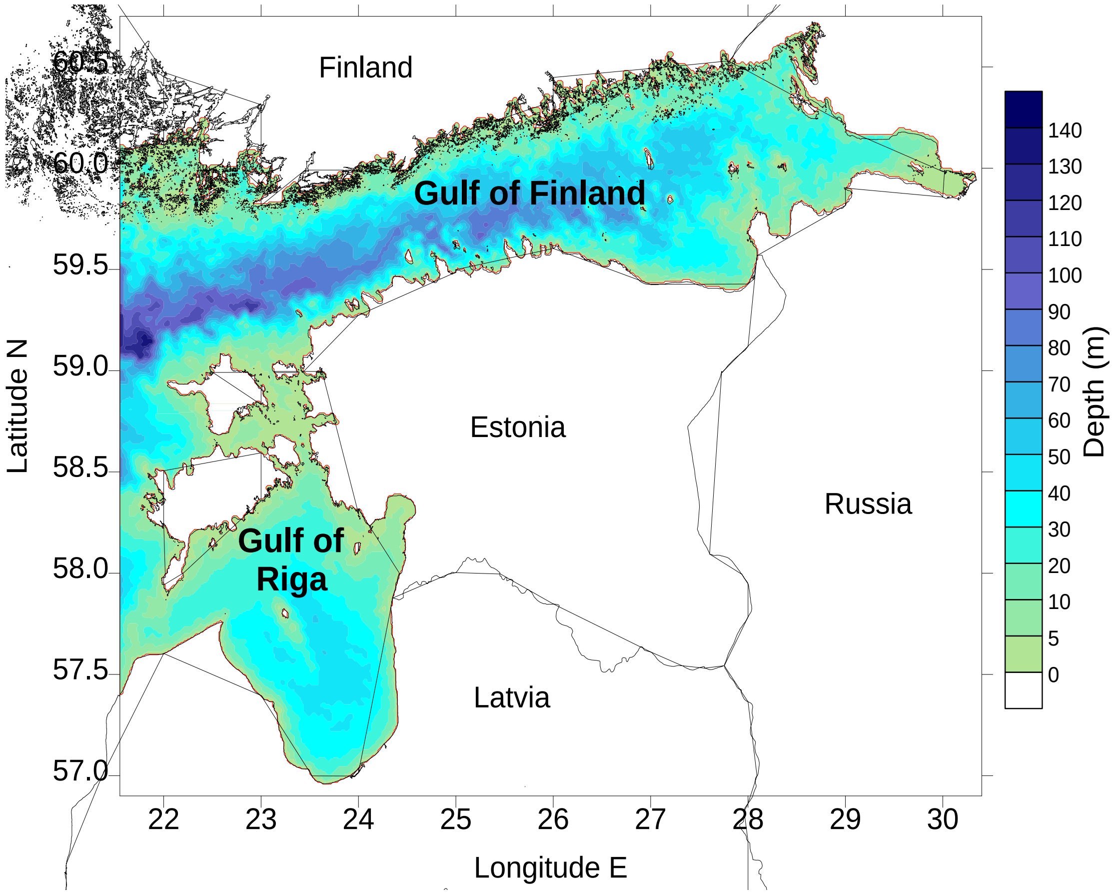
Methods (5): Experiments with pseudo-observations

Accuracy of EOF interpolation was checked by a series of experiments:
(a) configuration of observation points was selected;
(b) **model values were extracted at observation points** (we name them pseudo-observations);
(c**) interpolated fields were calculated** from pseudo-observations;
(d) calculations were **repeated for all time instances available**, statistical characteristics like RMSD between the interpolated and original fields were evaluated.

For comparison of the error with the spatial variability, RMSD of individual time instances were scaled with spatial standard deviations of the initial field.

We used following observation point configurations:
a) observations on **grid**, with step N times larger than model grid step, N = 1...8
b) typical **FerryBox** observation points
c) typical **monitoring** with reduced sampling network

Methods (6): Region of study and model



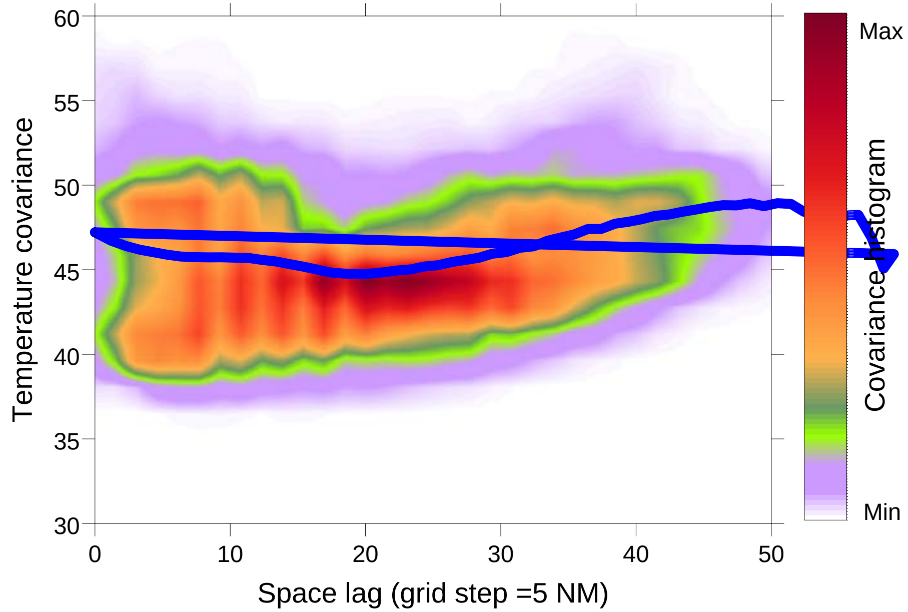
A map of the study area in the northeastern Baltic.

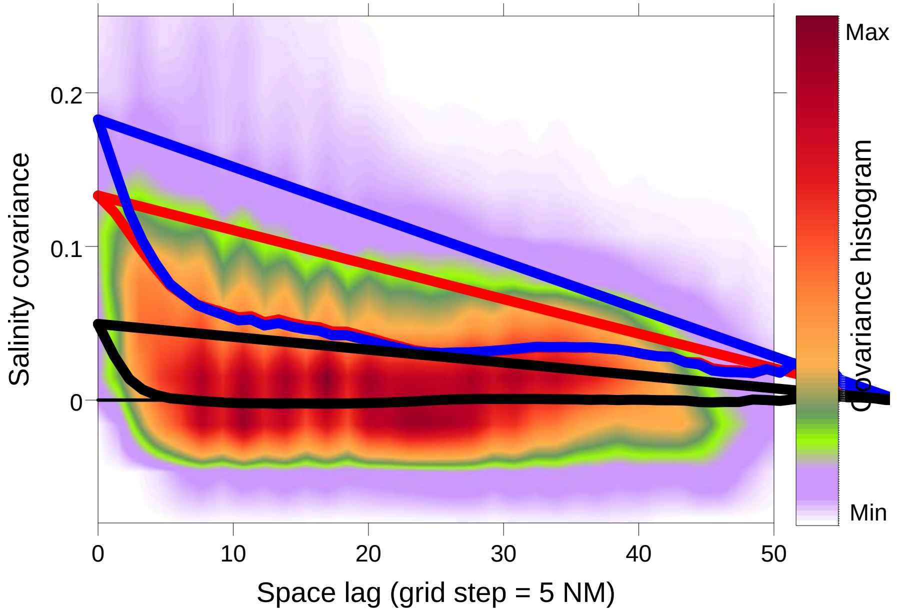
Oceanographic forecast model

HBM model with sub-regional 0.5 NM (nautical mile) setup in the geographical bounds shown in Fig to produce the SST and SSS data.

- Analysis of daily model data of free run (without data assimilation)
- **averaged over 10 x 10 grid points**, resulting 744 wet points with 5 NM resolution
- 5-year analysis period covered 1826 dates from July 1, 2010 to June 30, 2015

Results (1) : Covariance



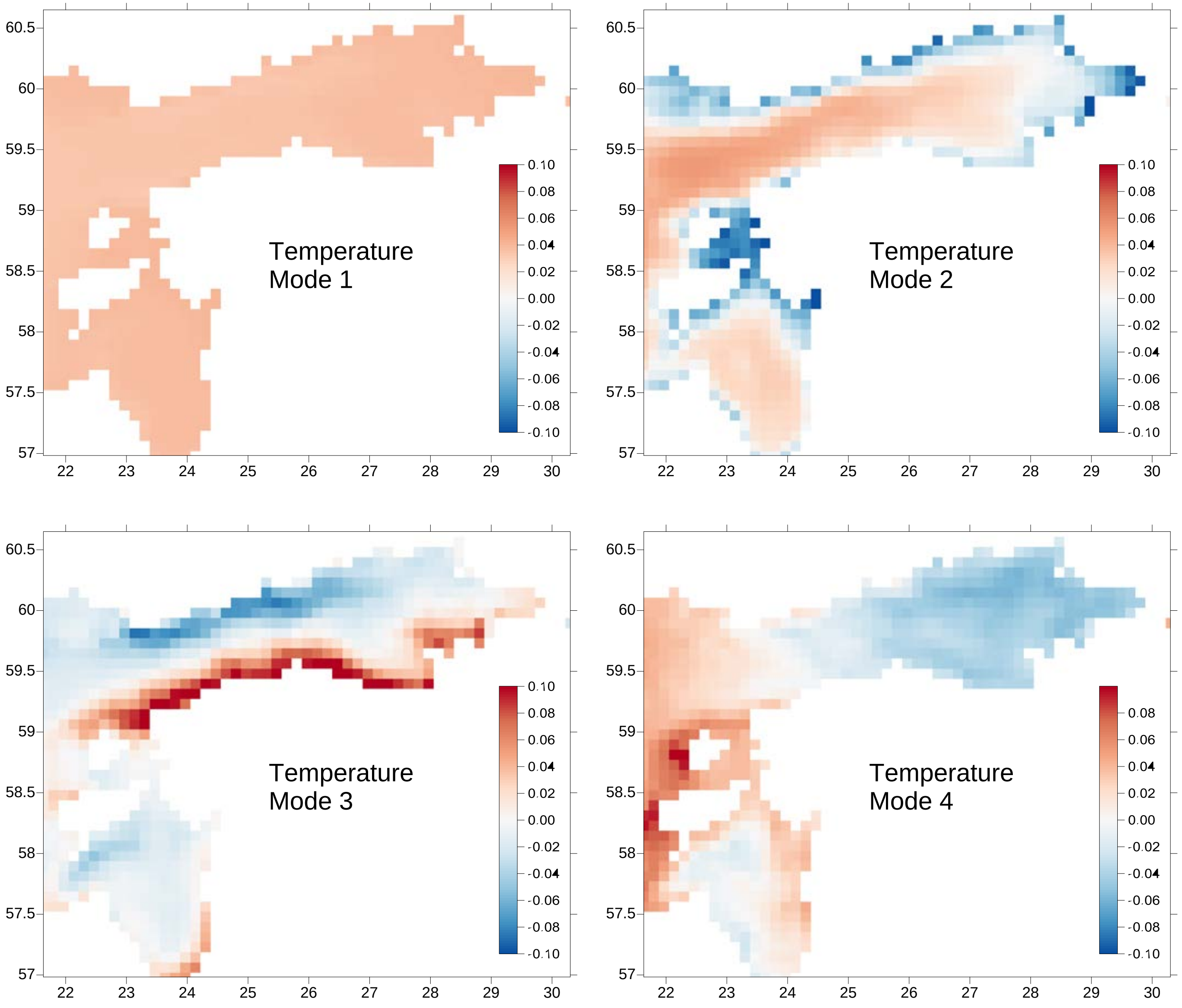


Covariance of SST (above) and SSS (below) **as a function of space lag** between the model points. Shown are heavily smoothed relative histograms of the original data (see color scale), mean covariance (blue line), covariance of six most energetic EOF modes (red line) and of higher EOF modes (black line) of SSS.

Distribution (histogram) of covariance in fixed space bins usually does not follow the normal distribution. Therefore, mean covariance values can be considered only as indicative.

Covariance of residual fields (sum of higher EOF modes) has a good normal distribution and it **decays fast with increasing space lag**: correlation goes below 0.2 at a distance of 16.5 NM (30 km) for both SST and SSS, justifying the use of localized interpolation methods for this part of the variability.

Results (2) : EOF mode patterns for SST



Mode 1 **97.64%**

Nearly uniform over space increase or decrease of SST, represents seasonal heating and cooling.

Mode 2 **1.28%**

Faster heating (in spring) or cooling (in autumn) in the shallow coastal areas, compared with deeper offshore areas.

Mode 3 **0.31%**

Transverse colder or warmer anomaly stripes near northern or southern coasts, like upwelling and downwelling.

Mode 4 **0.14%**

Longitudinal colder or warmer anomalies appearing in east-west direction.

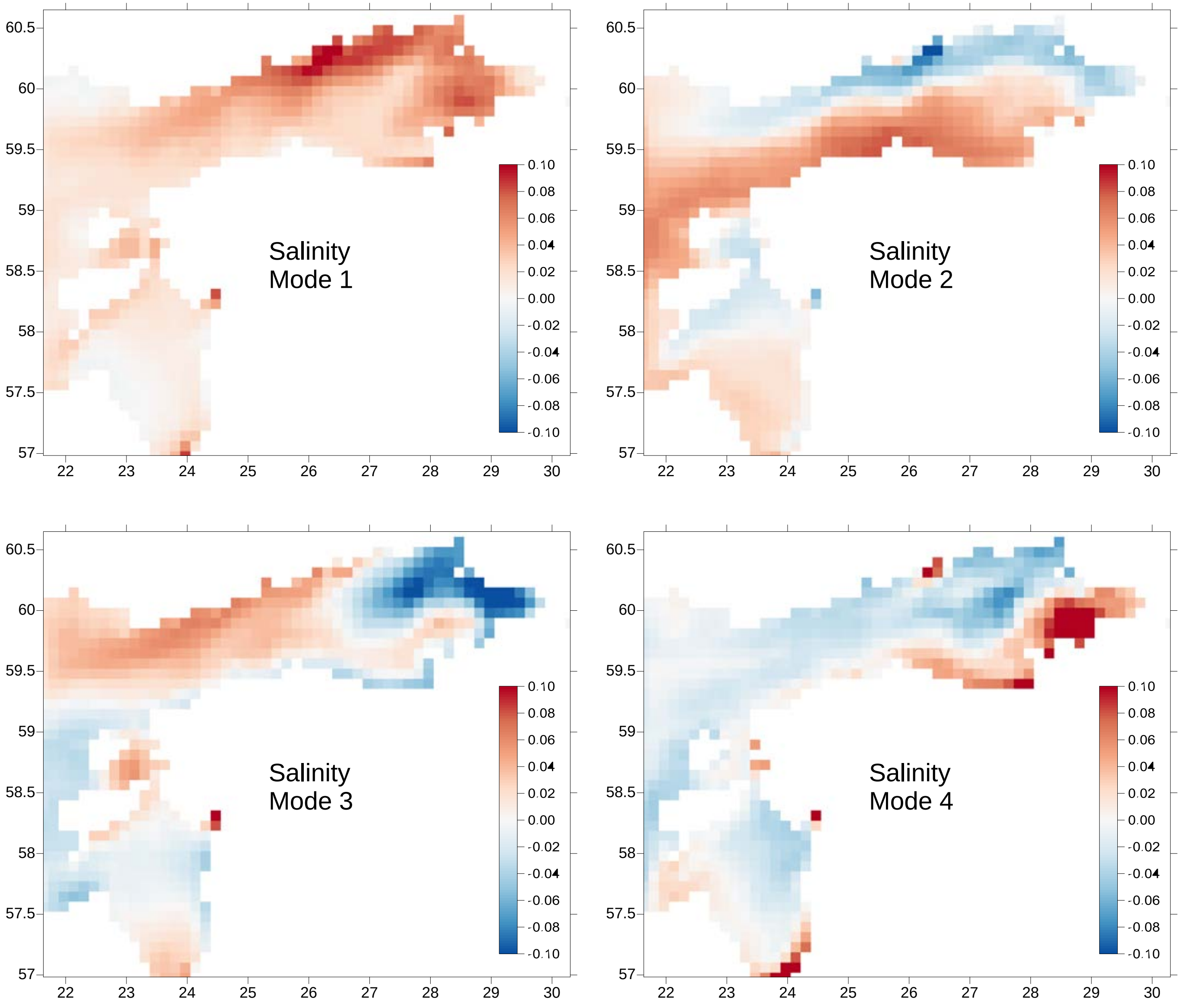
Mode 5 **0.10%**

Different heating or cooling of the SW Gulf of Riga and NW-N Gulf of Finland.

Mode 6 **0.07%**

Physics not clear.

Results (3) : EOF mode patterns for SSS



Mode 1 **36.17%**

Increase or decrease of salinity over the whole study area (all changes have the same sign). Larger changes occur in the northeastern Gulf of Finland, near the discharge of the largest rivers in the region.

Mode 2 **16.85%**

Transverse anomalies of salinity near northern or southern coasts, like upwelling and downwelling.

Mode 3 **7.06%**

Salinity changes in the freshwater spreading pathway near the northern coast of the Gulf of Finland, reminds cyclonic circulation.

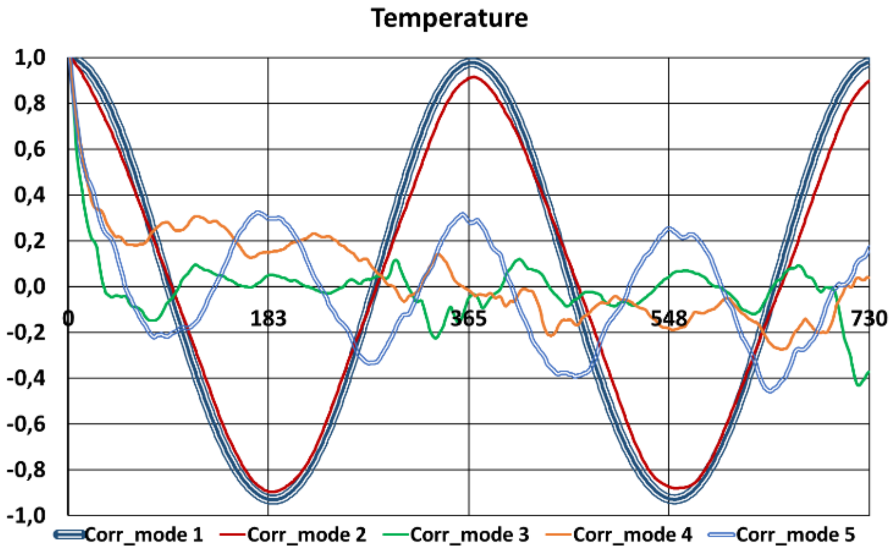
Mode 4 **5.17%**

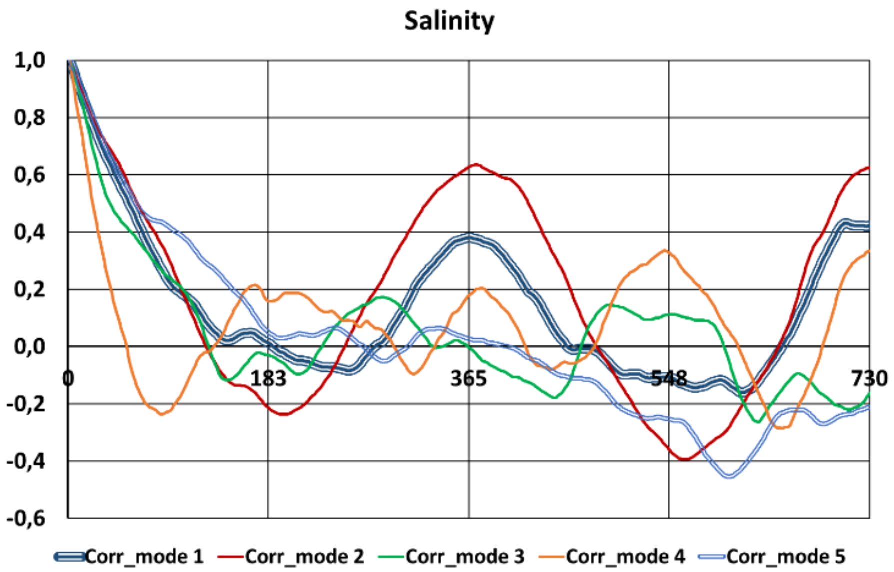
Salinity changes near the southeastern coasts, reminds anticyclonic circulation.

Mode 5 **4.11%**

Physics not clear.

Results (4) : EOF amplitudes



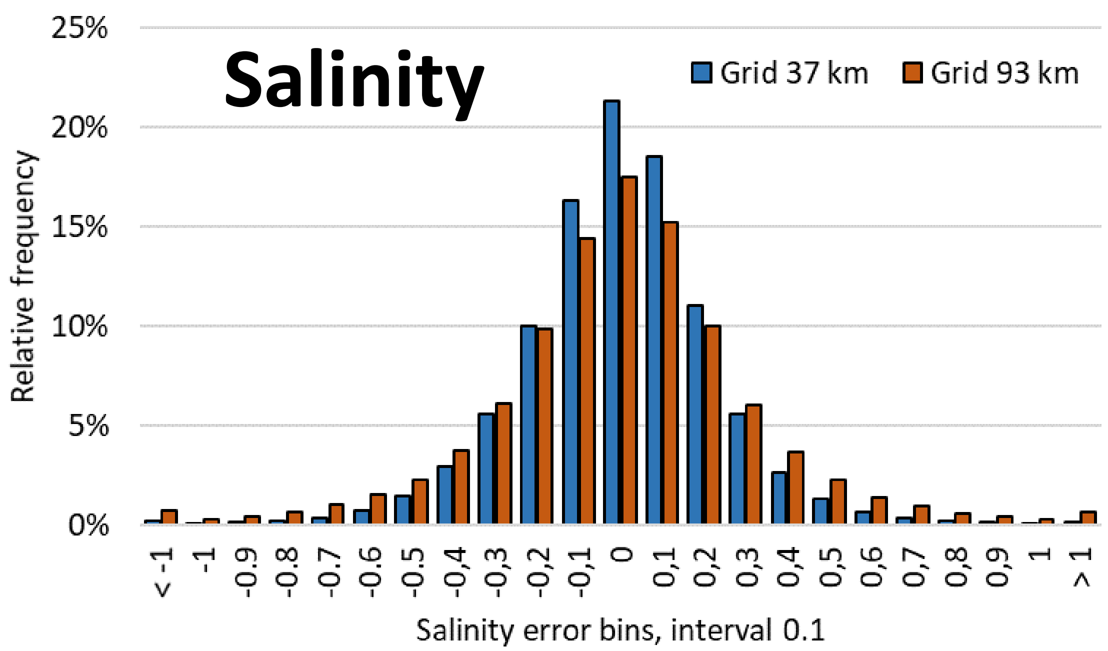


Temporal **correlation functions of EOF mode amplitudes** for SST (above) and SSS (below). Horizontal axis shows time lag in days.

Actual spatial observations are quite often not instantaneous in time. The weights of observations from past and future times depend on the temporal covariances (or correlations).

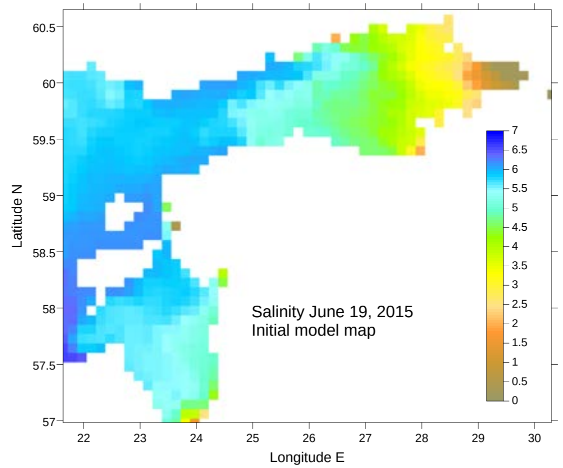
- Within the EOF decomposition, **amplitudes of SST and SSS modes have different temporal correlation patterns**.
- First and second SST modes are nearly annually periodic ($r > 0.9$), with shifted phases.
- First SSS mode has annual harmonic with $r \approx 0.4$. The second SSS mode has even stronger annual harmonic with $r \approx 0.6$.

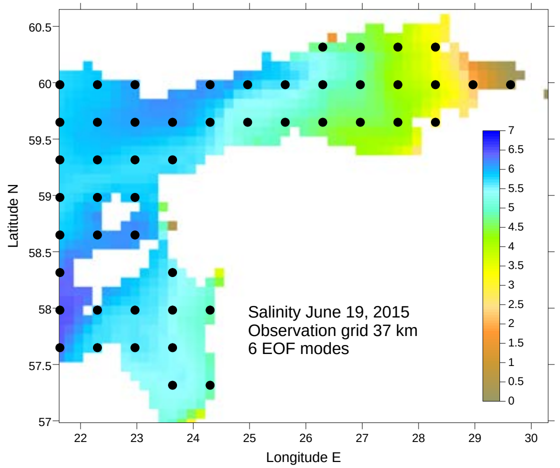
Results (5) : Interpolation errors, statistics



Examples of grid configuration for pseudo-observations. Correlation of initial and interpolated data is above 0.9.

Maps shown for SSS in 19.06.2015.

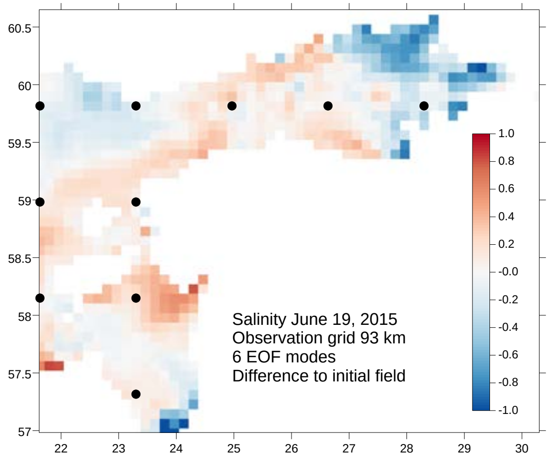


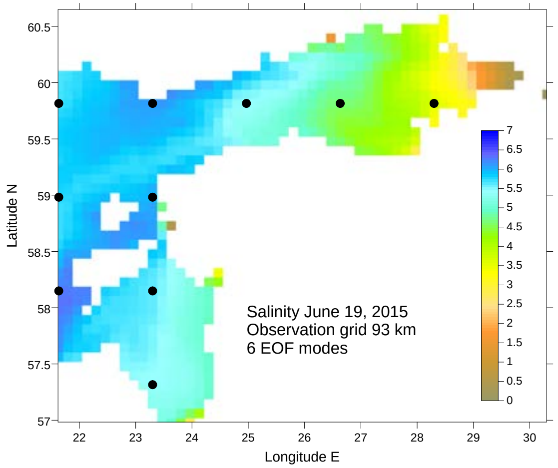


Relative frequency of SSS differences between

- interpolated (sum of six EOF modes) field &
- initial field.

Shown are results with pseudo-observation data prescribed by 37 km grid step (51 observation points) and 93 km grid step (10 points). Compared are about million data pairs.



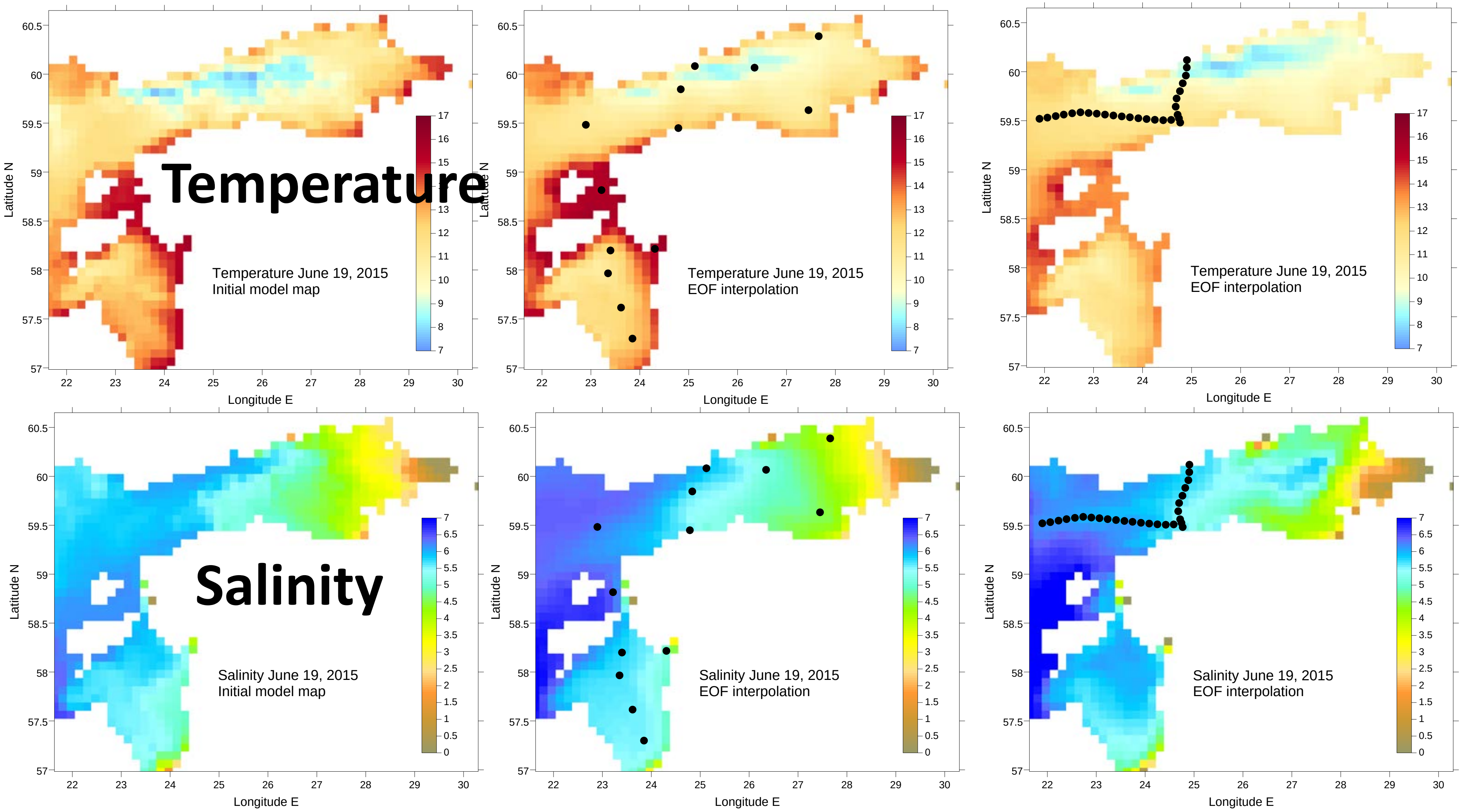


Results (6) : Interpolation errors, examples

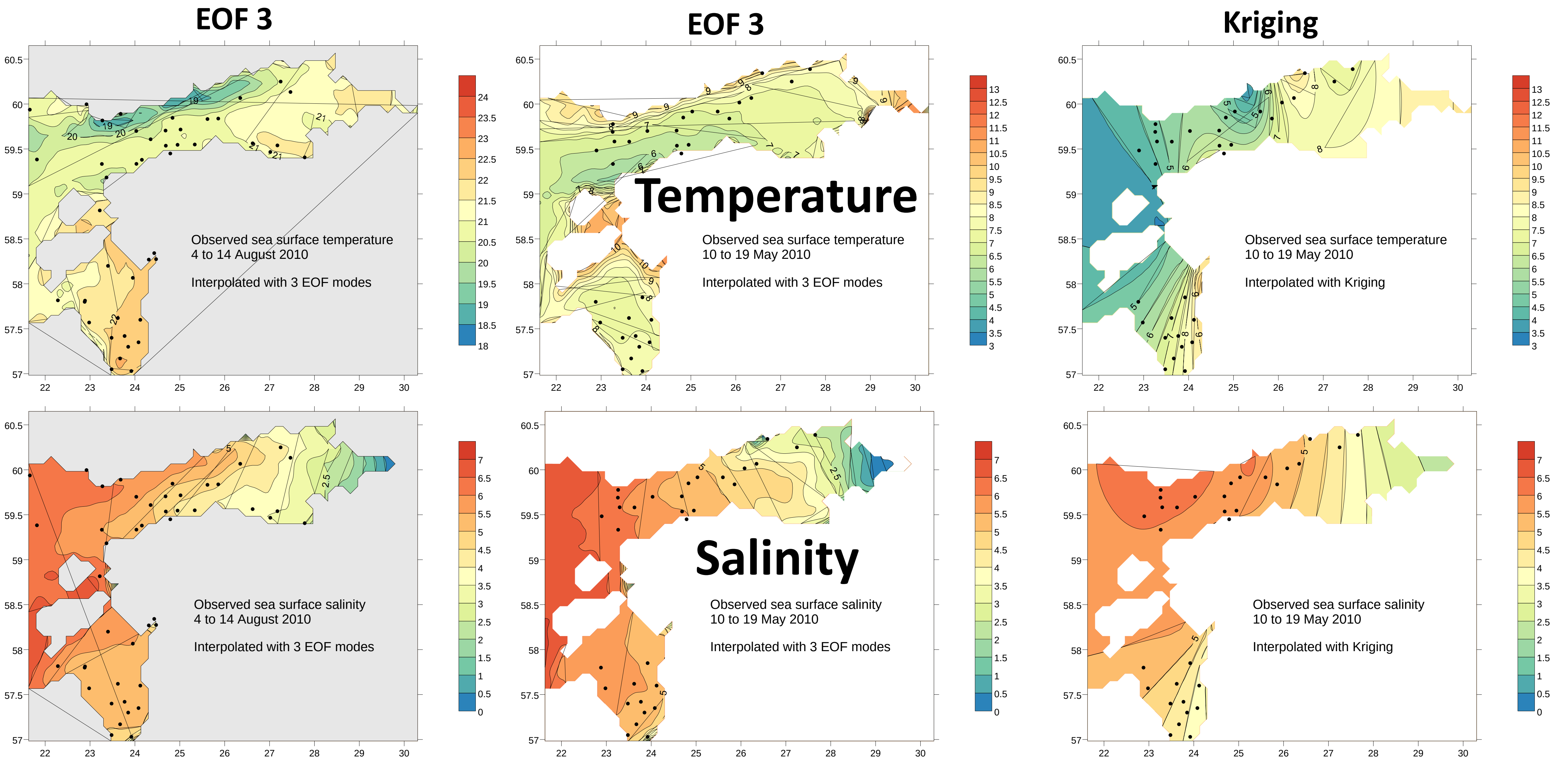
Initial model data

Interpolation from pseudo-observations: monitoring

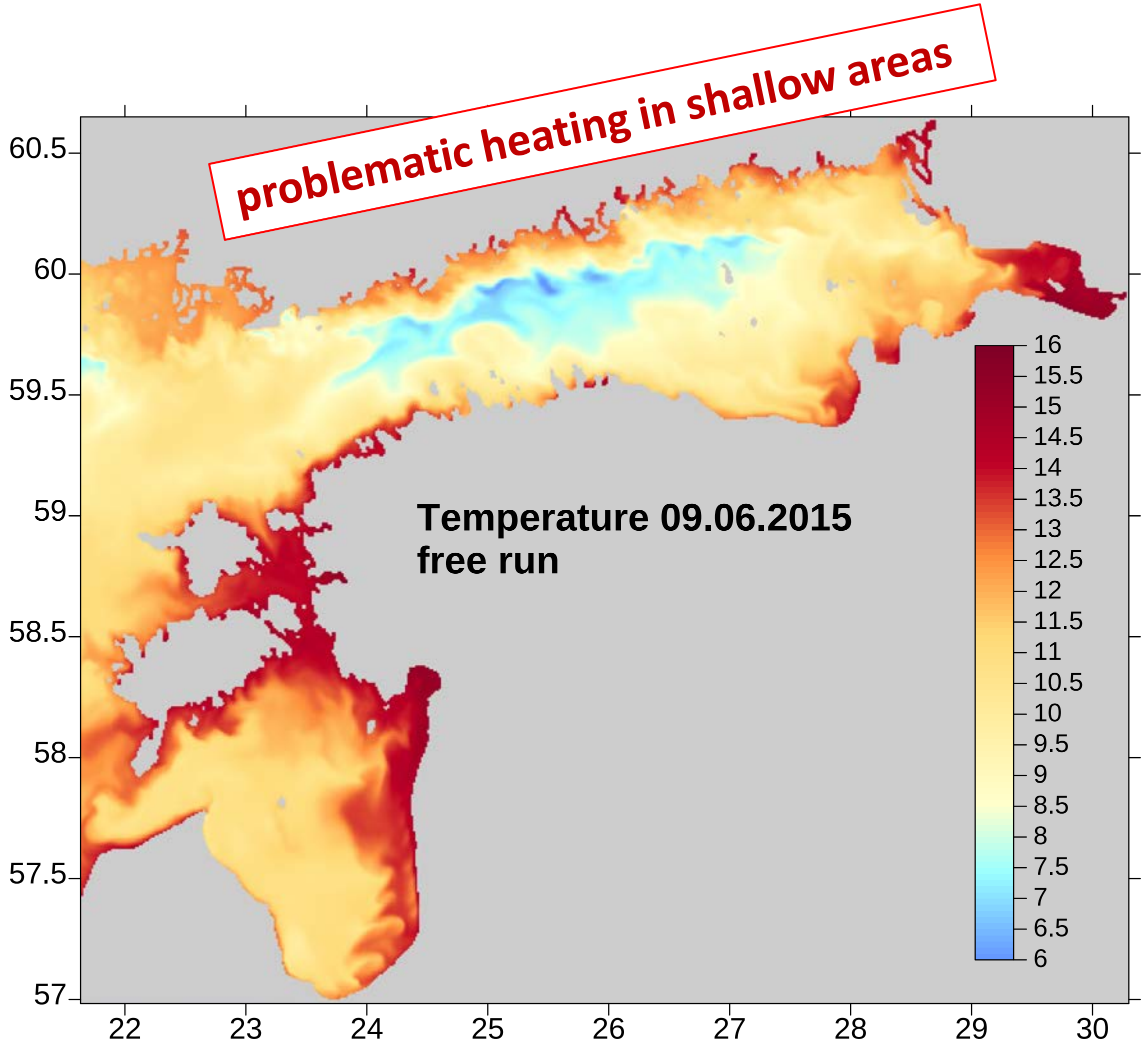
Interpolation from pseudo-observations: FerryBox



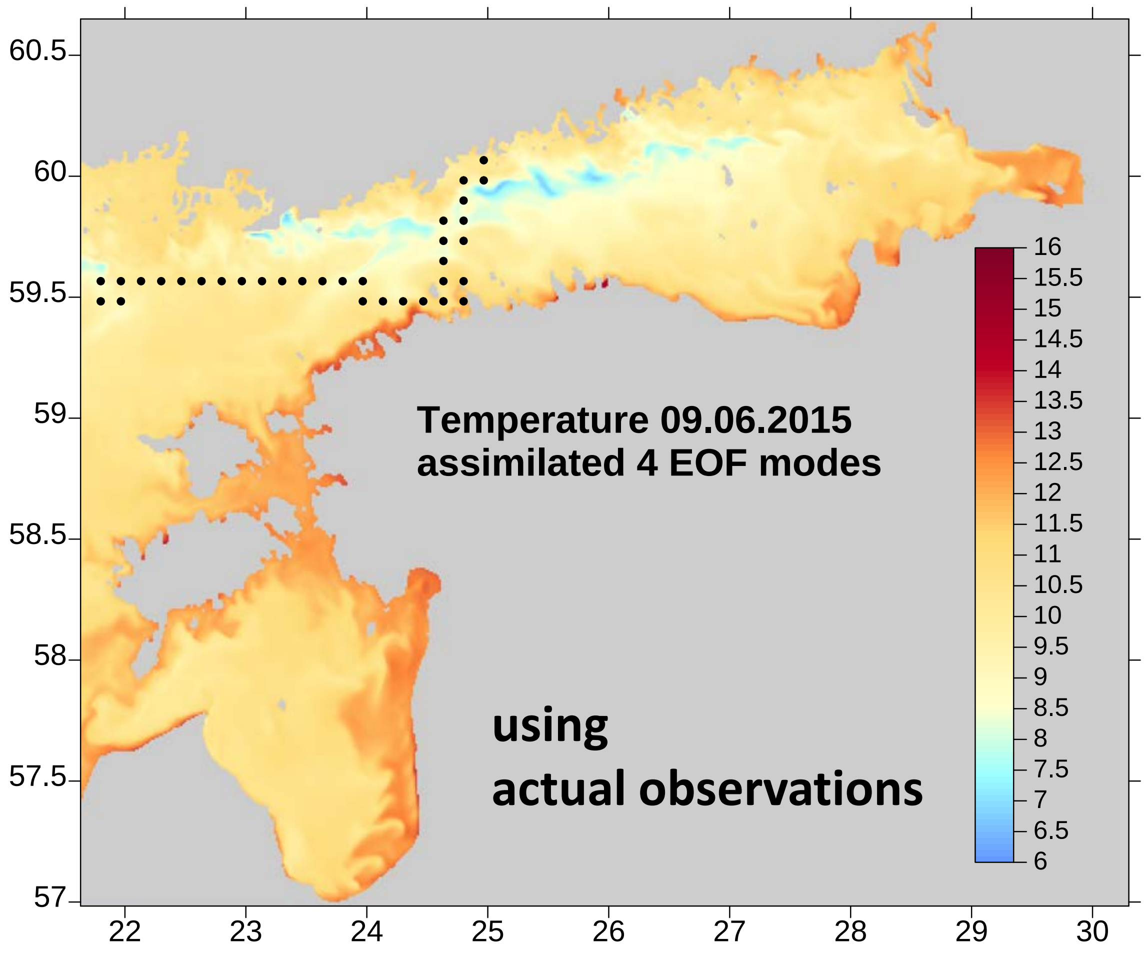
Results (7) : Interpolation of actual observations



Results (8) : One-step assimilation of FerryBox SST



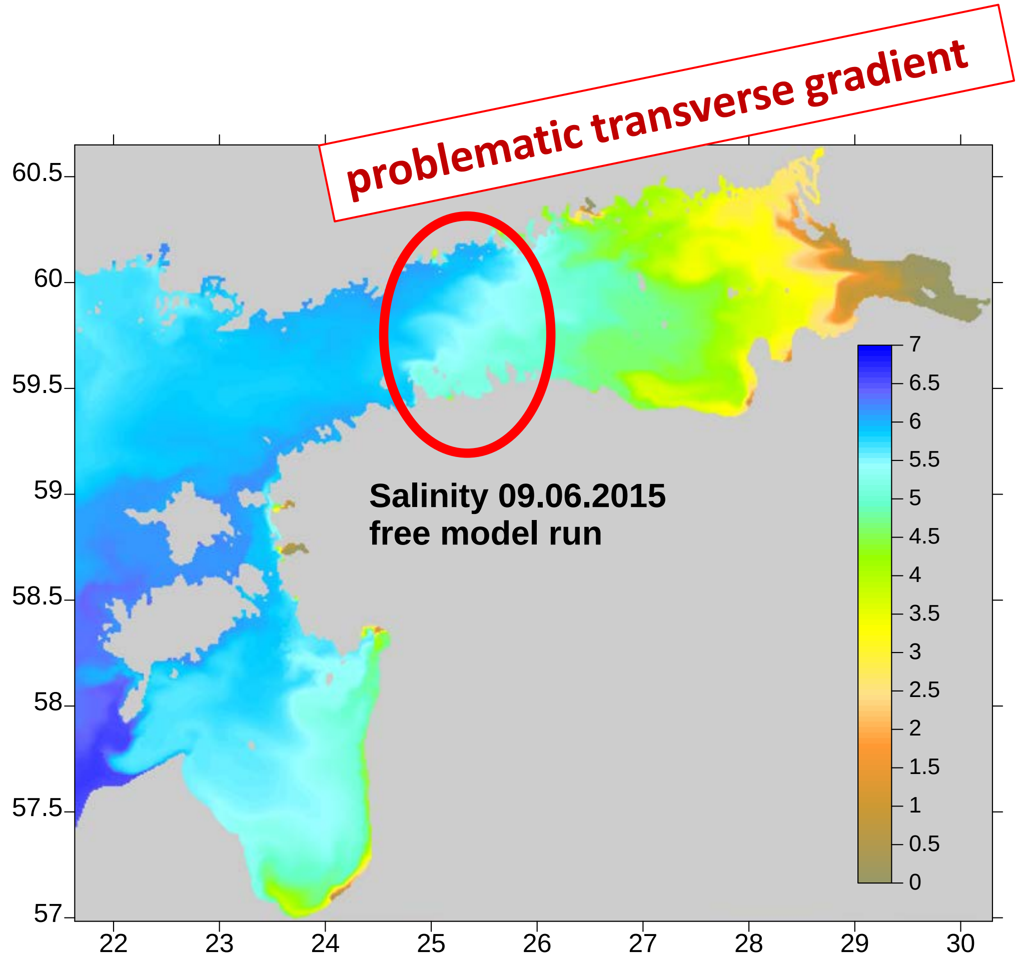
Initial field



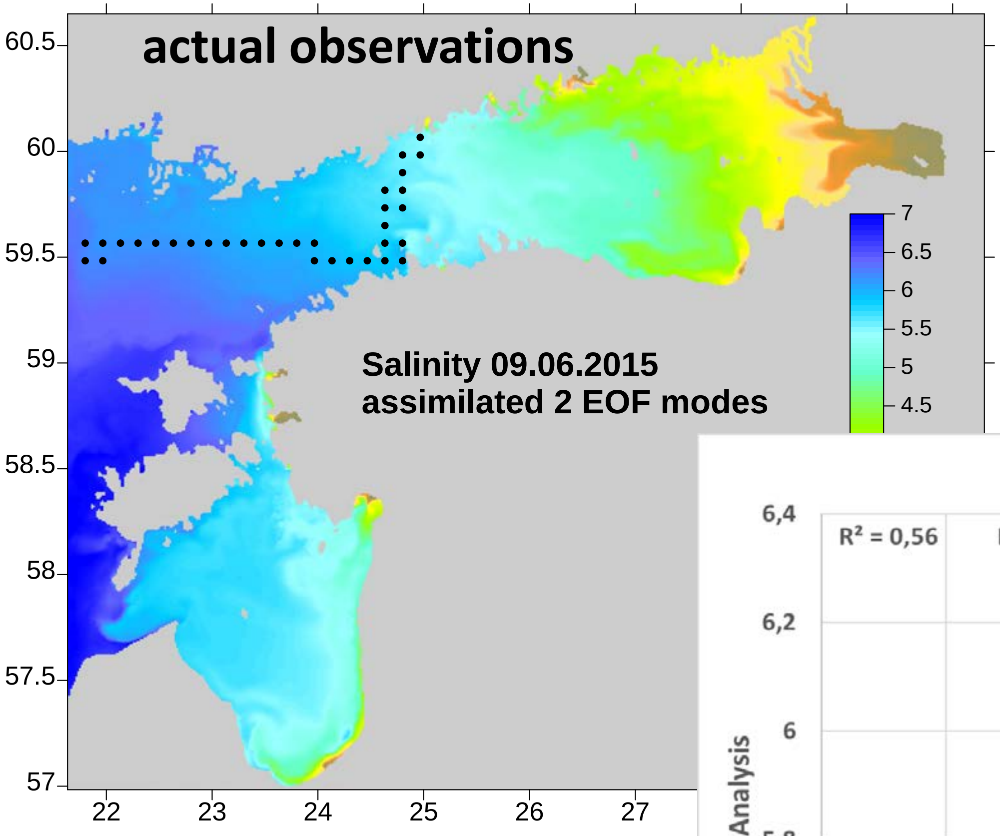
Assimilated field =
= 50% initial field + 50% EOF-interpolated field
Amplitudes of 4 EOF modes all are within limits

Coarse scale is corrected, fine-scale deviations remain the same.

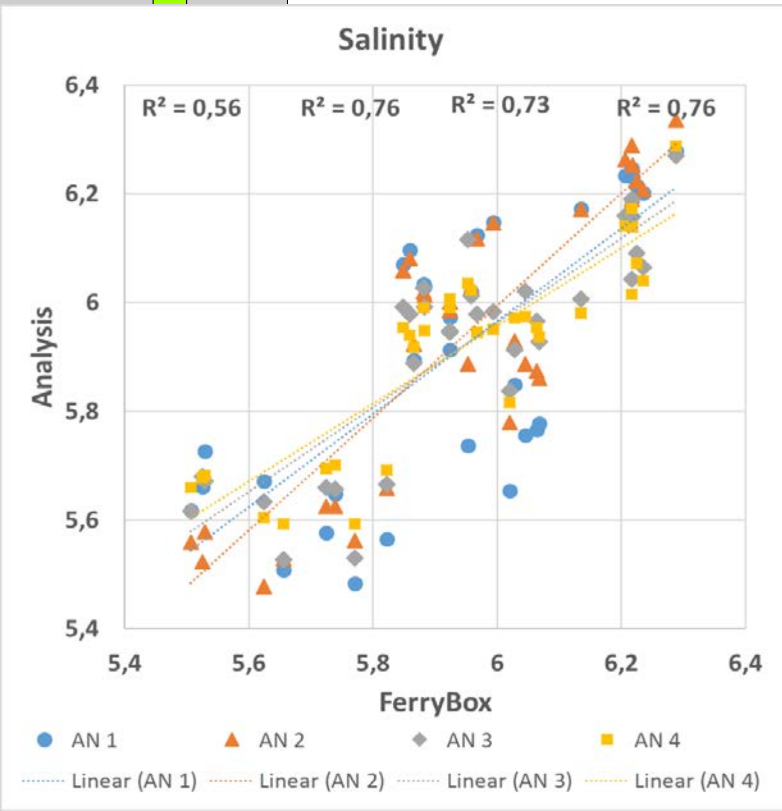
Results (9) : One-step assimilation of FerryBox SSS



Initial field

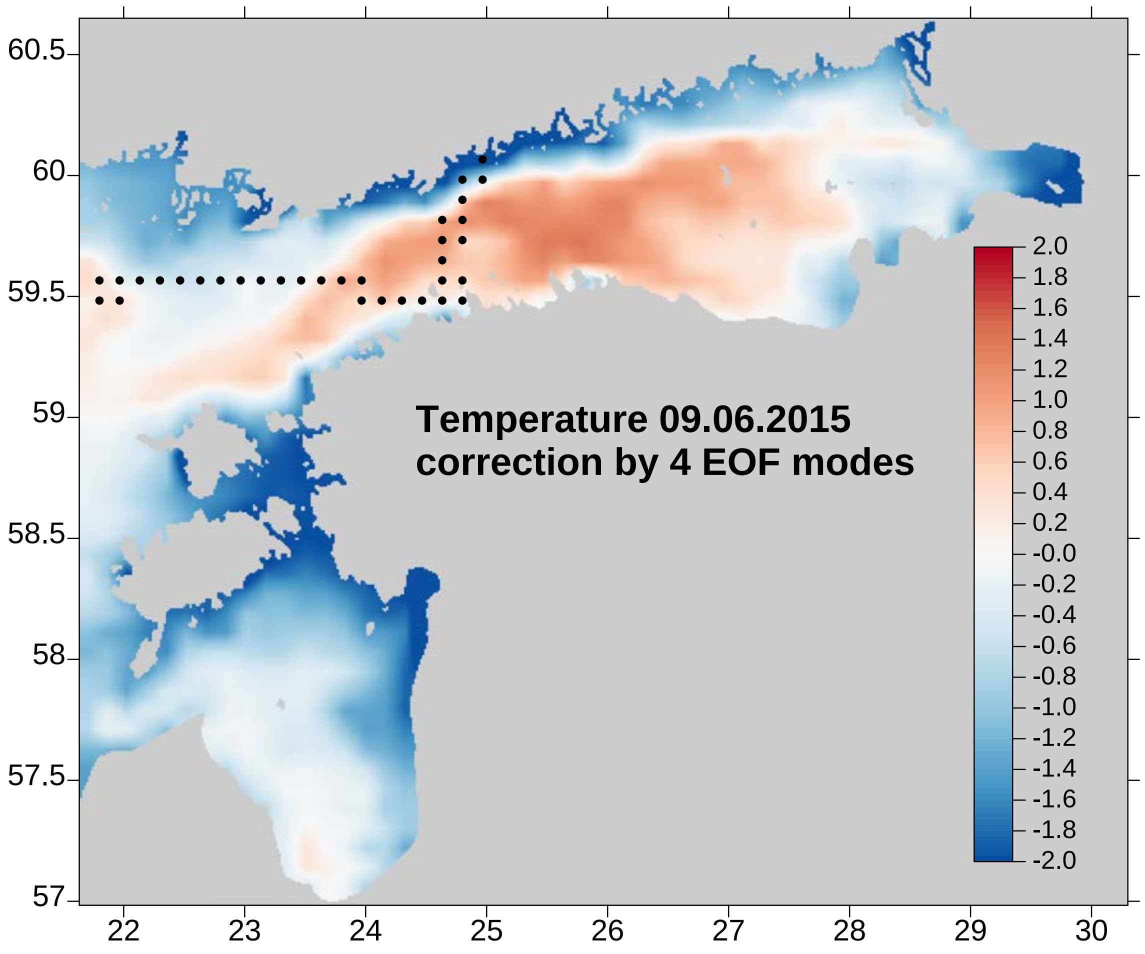


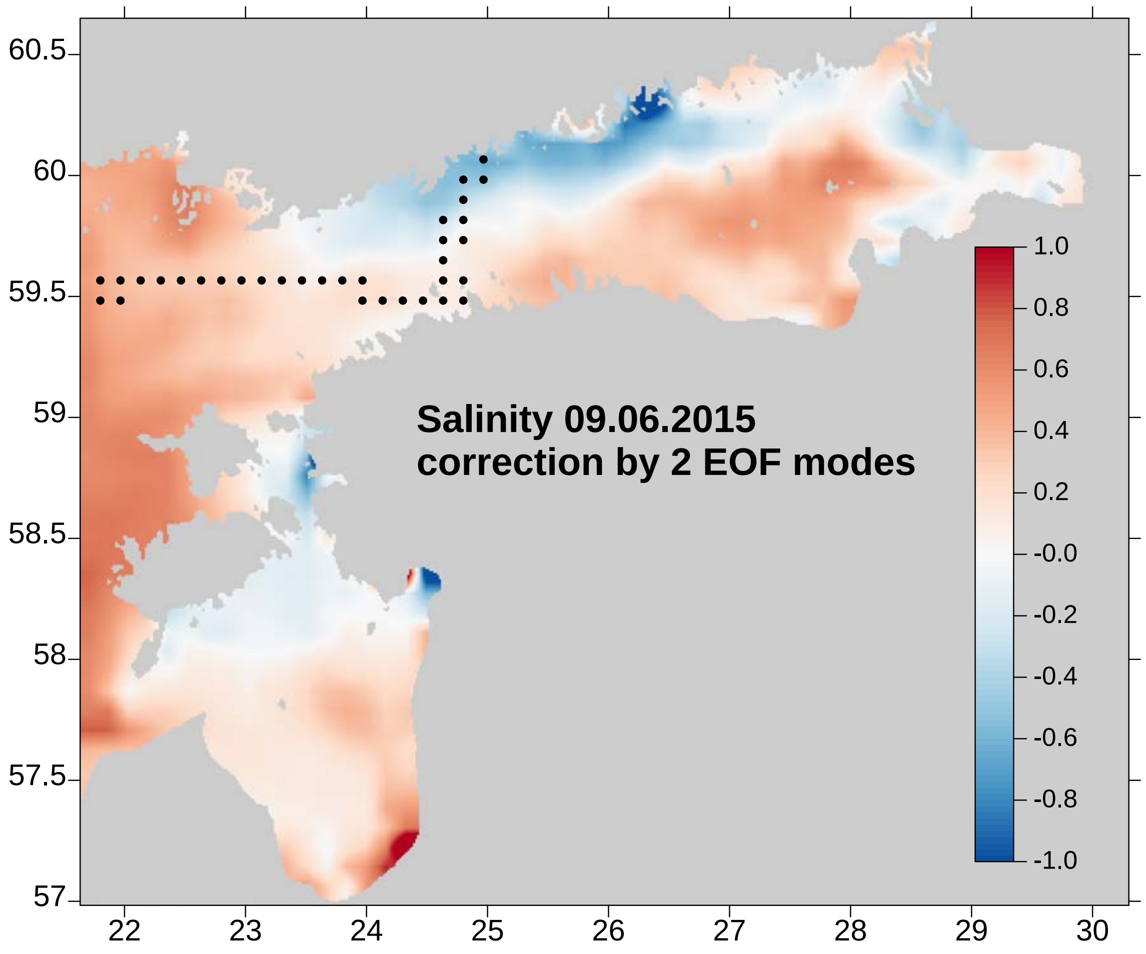
Initial model field is not correlated with FB observations. Assimilation introduces correlated pattern.



Assimilated field =
= 25% initial field +
+ 75% EOF-interpolated field
Amplitudes of only 2 EOF modes all are within limits

Results (10) : One-step corrections from FerryBox SST and SSS





Conclusions

- EOF interpolation is **statistically justified and computationally very effective** method to handle large-scale patterns in the sub-regions.
- In the smaller sea regions, which are **affected by the same large-scale forcing patterns**, the EOF patterns have **obvious physical interpretations** and their shape does not depend very much on the selection of boundaries.
- Since we use only the first most energetic EOF modes, we can cover with this method **basin and sub-basin scales of variability**.
- The **relative interpolation errors**, estimated over the full area, usually **are small** and remain below 10% for SST and 20% for SSS, compared with multi-year standard deviation of all variability relative to their mean value over the basin.
- **In the regions of denser sampling, EOF cannot follow the observations**. Mesoscale deviations from large-scale EOF patterns follow well-defined covariance decay with space lag; therefore they could be treated by optimal interpolation or similar method.
- The developed EOF interpolation can serve as a **first simple and robust step in data assimilation**.