The maximum entropy approach for a posteriori estimation of model and data errors

Svetlana Losa

Alfred Wegener Institute for Polar and Marine Research

Bremerhaven, Germany

Thanks to Gennady Kivman, Sergey Danilov, Frank Janssen, Jens Schröter, Vladimir Ryabchenko, Tijana Janjić

Data assimilation in oceanography

(state x_i and parameter estimation)

Dynamical model $\qquad L_p(x) = F, \qquad$ L is model operator

uncertainies in nitial condition $x(0)$, model parameters p, external forcing F

$$\rho_t^f(x(t), p) = C\rho^f(x(t) \mid x(0), p)\rho(p)\rho_0(x(0))$$

defined on a $X_t \times P$ space, $x(t) \in X_t$, $p \in P$

Observational data $\qquad H(x) = d, \qquad$ H is observational operator

$$\rho(x, p \mid d) = C\rho(d \mid x)\rho(x, p)$$

$$\rho(x, p) = C\rho(p)\rho_0(x(0))\prod_{k=1}^{M}\rho^f(x(k\delta t) \mid x((k-1)\delta t), p)$$

We are not confident about the model and data uncertainties.

Do we need the uncertainties quantification?

Principle of Maximum Entropy

(general formulation, Kivman et al., 2001)

$$He(\rho) = -\int_X \rho(x \mid d) \ln \frac{\rho(x \mid d)}{\mu(x)} \prod dx$$

$\mu(x)$ is the lowest information about x.

The maximum probable x or mean with respect to $\rho(x|d)$ is

$$x_i = M_m x_m + M_d x_d$$

$$M_m = L_* L, \quad M_d = H_* H$$

L_*, H_* reflect our assumptions on the model and data error covariances.

Operators M_m and M_d are nonnegative, self-adjoint and

$$M_m + M_d = I$$

$$(x_i) = Arg\min\left\{\int_0^T [L(x,t) - F(t)]^2 dt + \beta \sum_{m=1}^{M} (H(x) - d)^2\right\}$$

$$He(M) = -trace(M_d \ln M_d + M_m \ln M_m) = -\sum_{i=1}^{N} [\lambda_i \ln \lambda_i + (1 - \lambda_i) \ln(1 - \lambda_i)]$$

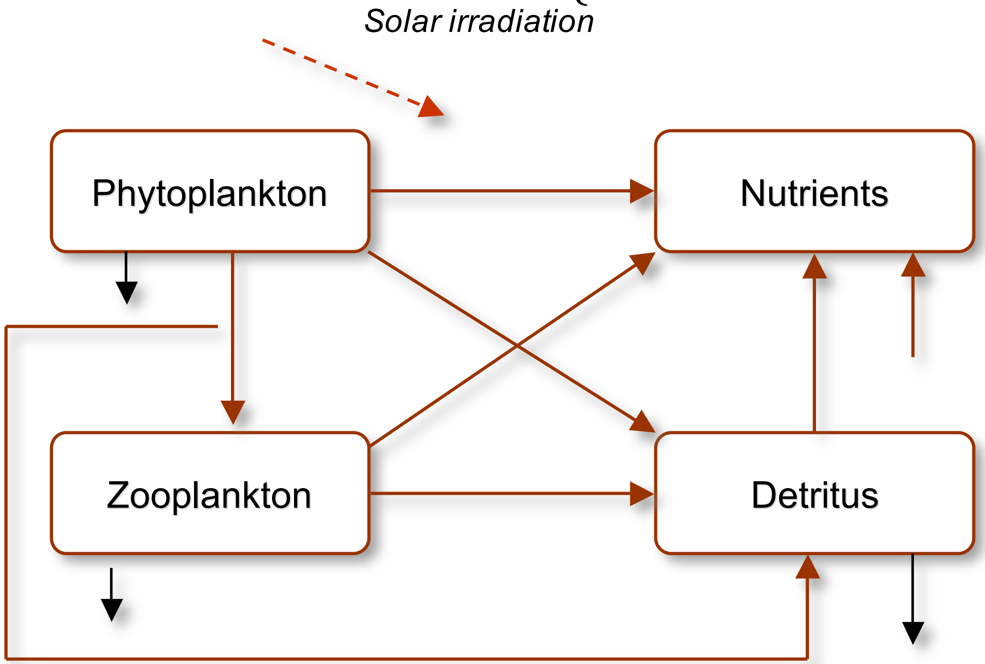
M is an operator-valued measure.

We have to find β which would maximize $He(M)$ … or correspondent term in the cost function (Maximum Data Cost, MDC)

Popova's Ecosystem Model (1995)

(generalized inversion)

$$(x,p) = Arg\min\left\{\int_0^T \left[\frac{dx}{dt} - L_p(x)\right]^2 dt + \beta\sum_{m=1}^{M}(H(x)-d)^2\right\}$$

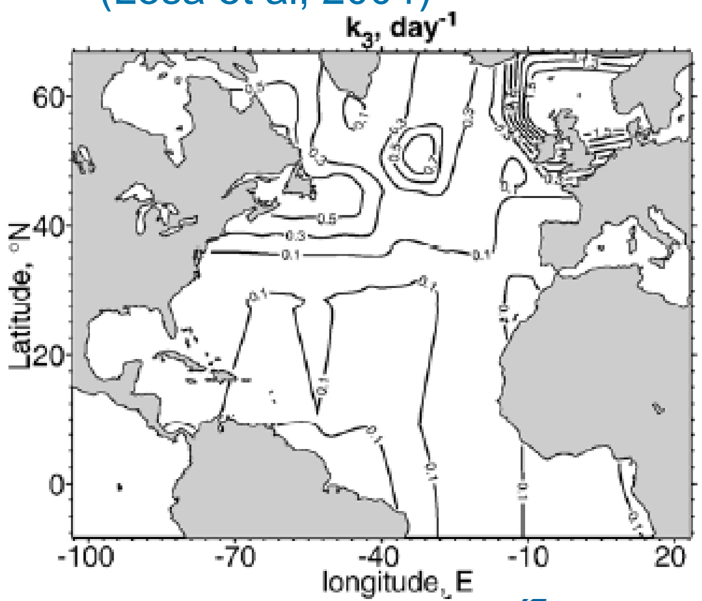


The flow network between 4 biogeochemical {P, Z, N, D} components, x, possesses 19 biological parameters, p.

6 of them have been adjusted for each cell of 5^0x5^0 grid covering the North Atlantic

Assimilated data:
Monthly mean satellite CZCS surface chlorophyll averaged over 1979 – 1985.

Method : a weak constraint variational technique (Losa et al, 2004)

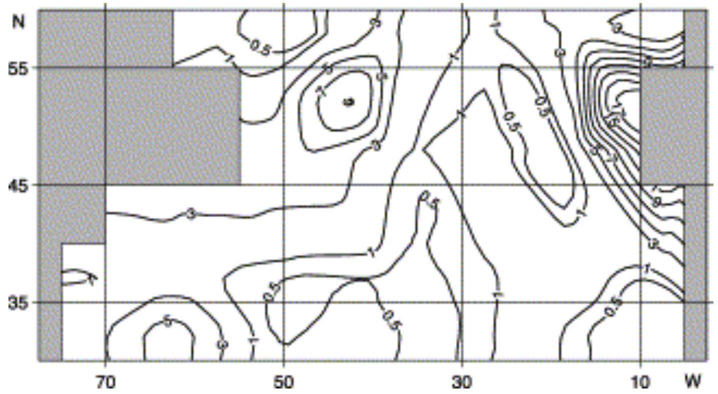


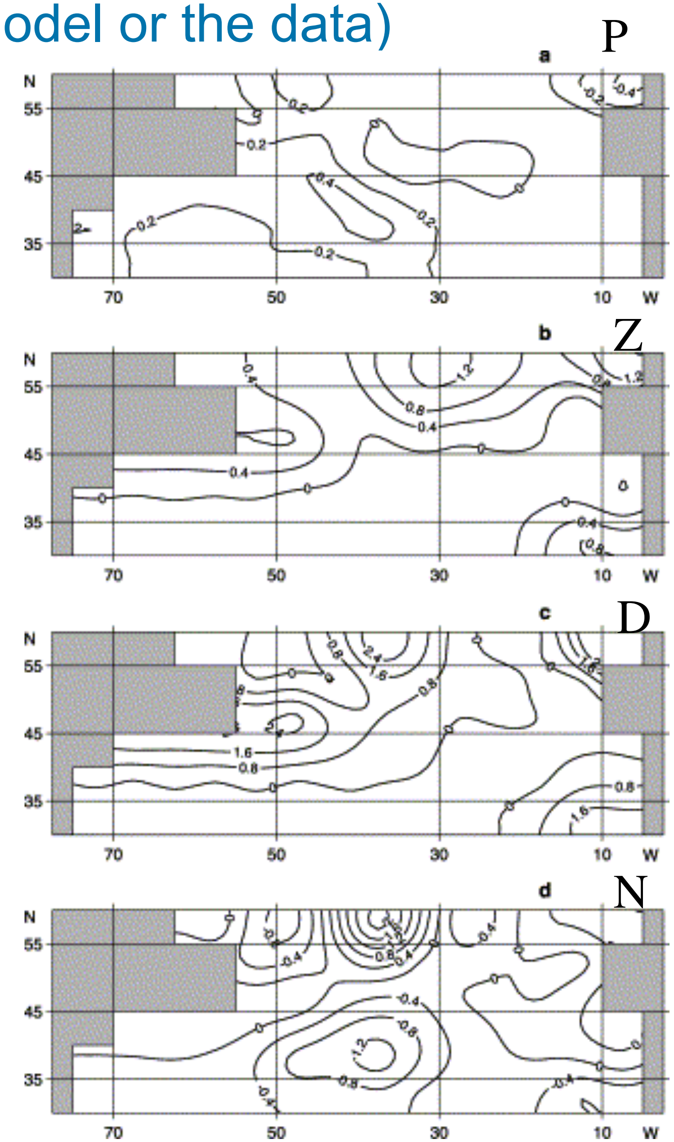
Inference about the model and the data

(Which is better: the model or the data)

The ratio of the terms in the cost function

$$r = \frac{\beta \sum_{m=1}^{M} \left| H(x) - d \right|^2}{\int_0^T \left[\frac{dx}{dt} - L_p(x) \right]^2 dt}$$





Annual model equation residuals normalized by the total biological source $\Rightarrow$

Inference about the model parameterizations and fluxes

(Secondary Inversion)

-- model solution

- inverse solution

▬ secondary inversion

$$\frac{d\tilde{x}}{dt} = L_p(x)$$

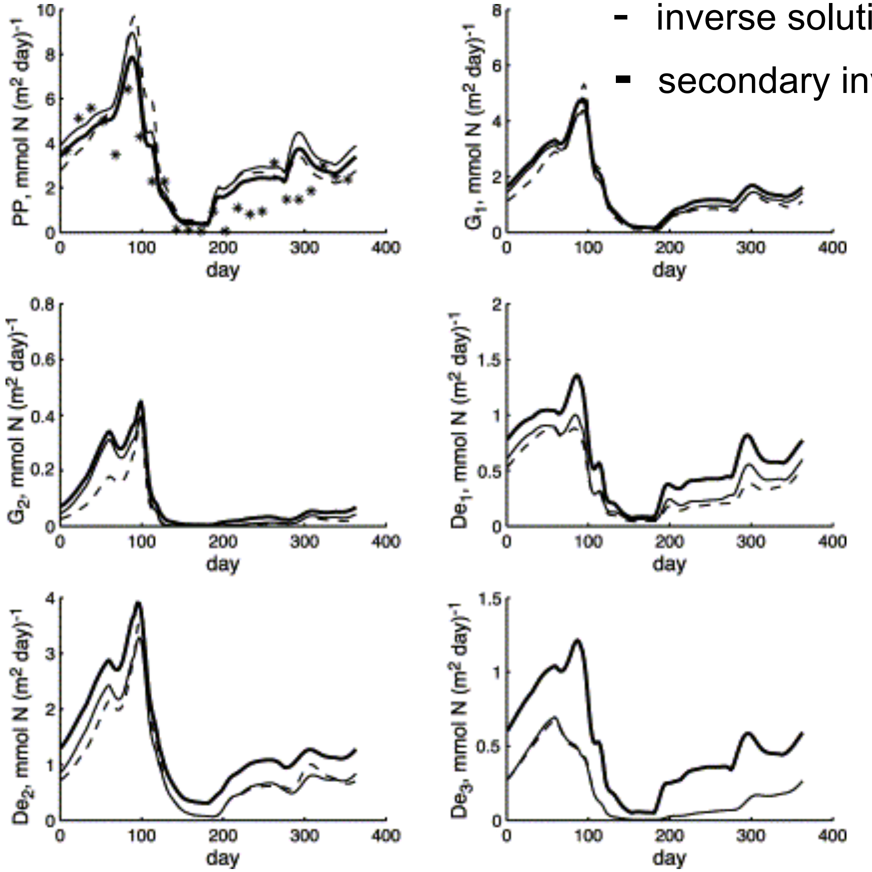
The first guess

$$P\tilde{P}, \tilde{G}_1, \tilde{G}_2, \tilde{D}e_i, i = 1,2,3$$

is calculated given

$$\tilde{x} = (\tilde{P}, \tilde{Z}, \tilde{D}, \tilde{N})$$

and optimal parameters

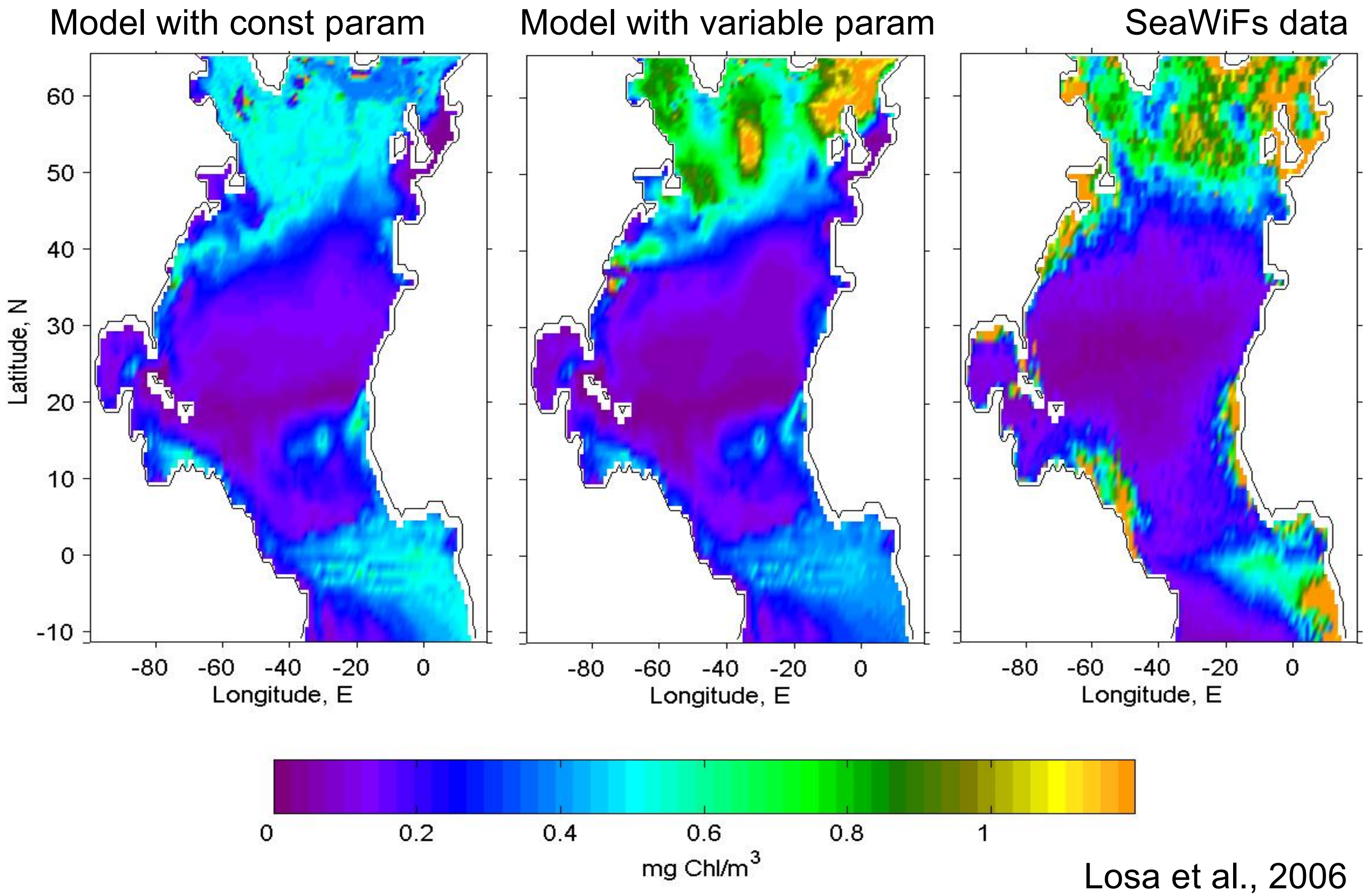


$$(PP, G_1, G_2, De_1, De_2, De_3) = Arg\min\left[(PP - P\tilde{P}) + \sum_{i=1}^{2}(G_i - \tilde{G}_i)^2 + \sum_{i=1}^{3}(De_i - \tilde{D}e_i)^2\right]$$

August horizontal distribution of the surface chlorophyll “a” in the North Atlantic

AWI

(Popova’s NPZD coupled to 3D POP gcm)

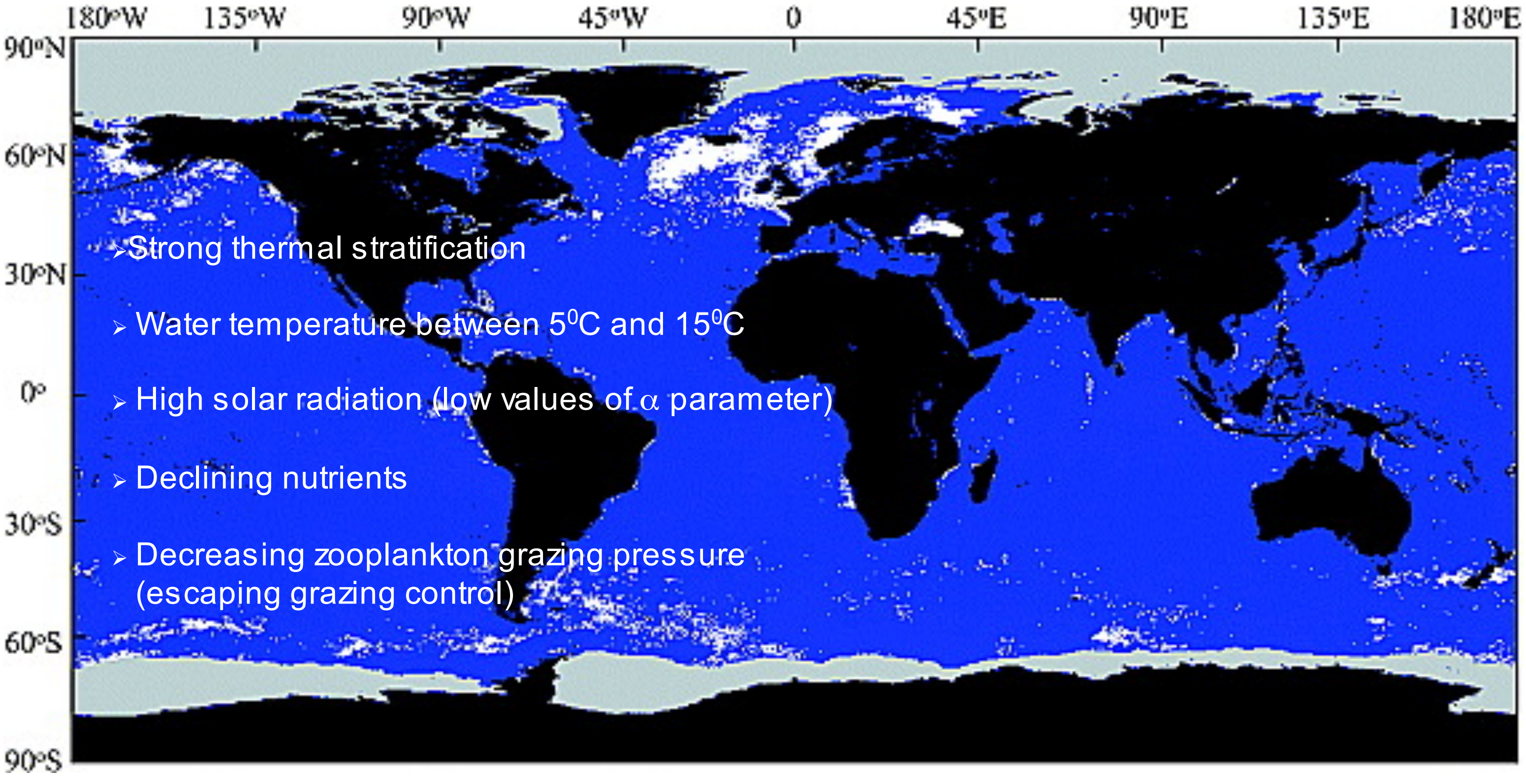


Losa et al., 2006

HELMHOLTZ GEMEINSCHAFT

Annual composite of classified coccolithophorid blooms in SeaWiFS imagery dating from October 1997 to September 1999

(Iglesias-Rodríguez et al., 2002)



- Strong thermal stratification
- Water temperature between 5^0C and 15^0C
- High solar radiation (low values of α parameter)
- Declining nutrients
- Decreasing zooplankton grazing pressure (escaping grazing control)

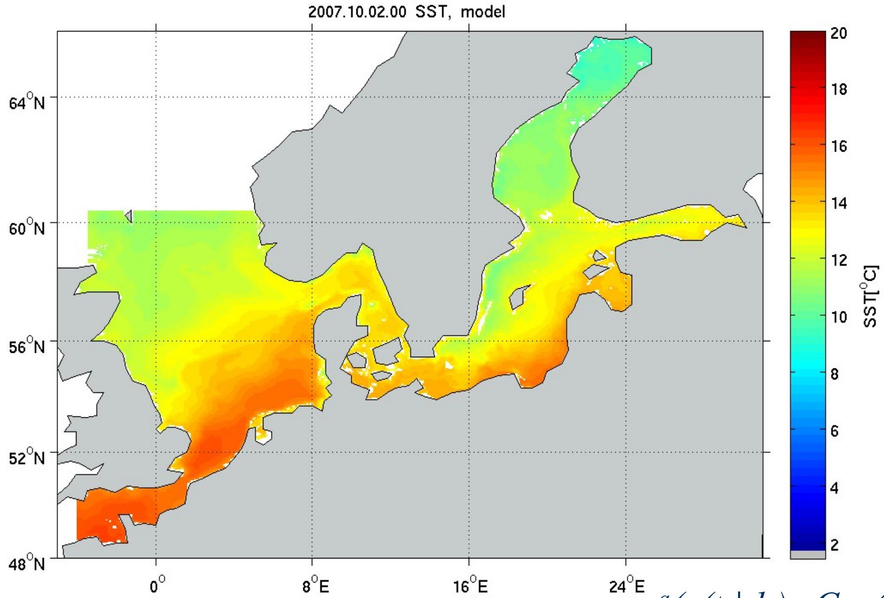
The bloom class is white, the non-coccolithophorid bloom class is blue, the land is black, and ice is gray.

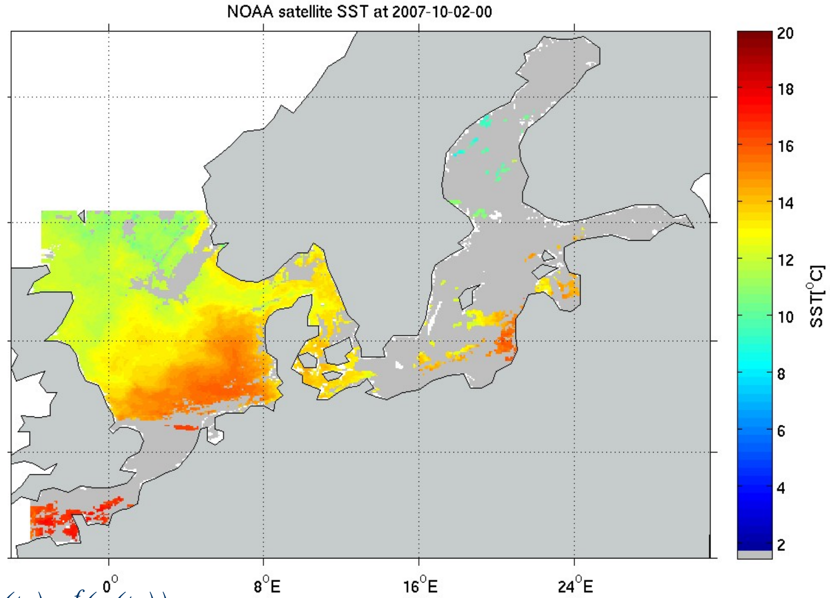
Assimilating NOAA's SST data into an operational circulation model of the North and Baltic Seas

AWI

BSHcmod *run at the*
German Maritime and Hydrographic Agency (BSH)

NOAA SST
12 hourly-around 00:00 and 12:00,- composites of SST measured by the Advanced Very High Resolution Radiometer (AVHRR) aboard polar orbiting satellites





$$\rho_t^a(x(t_1|d_1)=C\rho_d(d_1|x(t_1)\rho_t^f(x(t_1))$$
$$\rho_t^f(x(t_1)=C\rho^f(x(t)|x(0))\rho_0(x(0))$$

Extraction and combination of the information from two different sources - the model and the data - in order to improve our understanding of both sources and, therefore, of reality itself

HELMHOLTZ
GEMEINSCHAFT

Sequantial statistical approach

(Kalman type filtering)

Ensemble based Singular Evolutive Interpolated Kalman filter (SEIK, Pham, 2001)

$$x(t_n)^a = x(t_n)^{f,m} + K_n(d_n - Hx(t_n)^{f,m})$$

$$K_n = P_n^f H(HP_n^f H^T + R)^{-1}$$

x^f, x^a denote forecast and analysis of state vector (at time t_n at all grid points)

d_n - observations available (at t_n)

P_n^f - forecast error covariance matrix

R - observational error covariance matrix

SEIK Filter is implemented locally (PDAF, Nerger et al., 2006) but with different formulations of data error correlation.

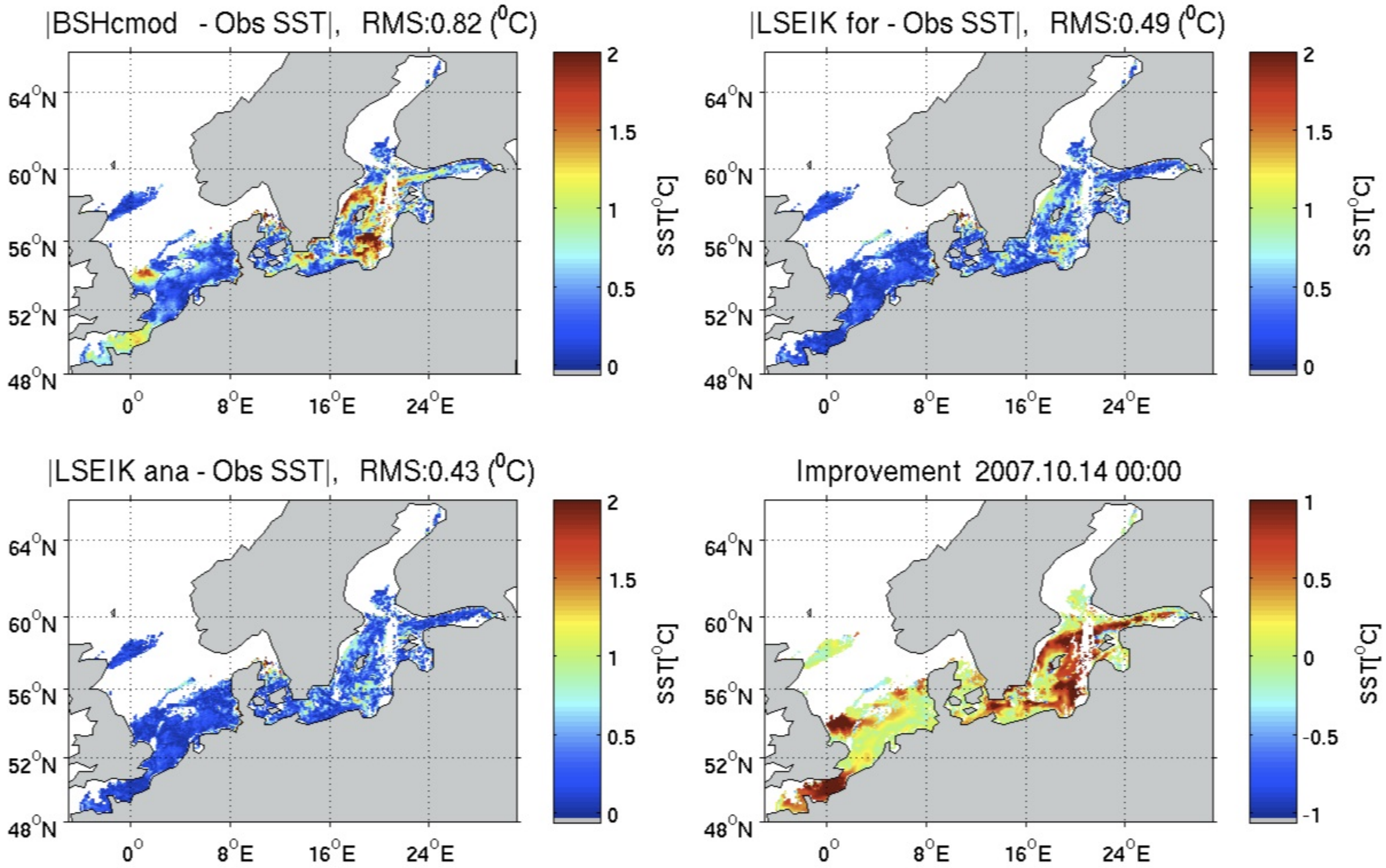
When calculating $He(M)$, the Kalman gain K could be considered

globally over a certain period of time

locally (for validation of localization conditions)

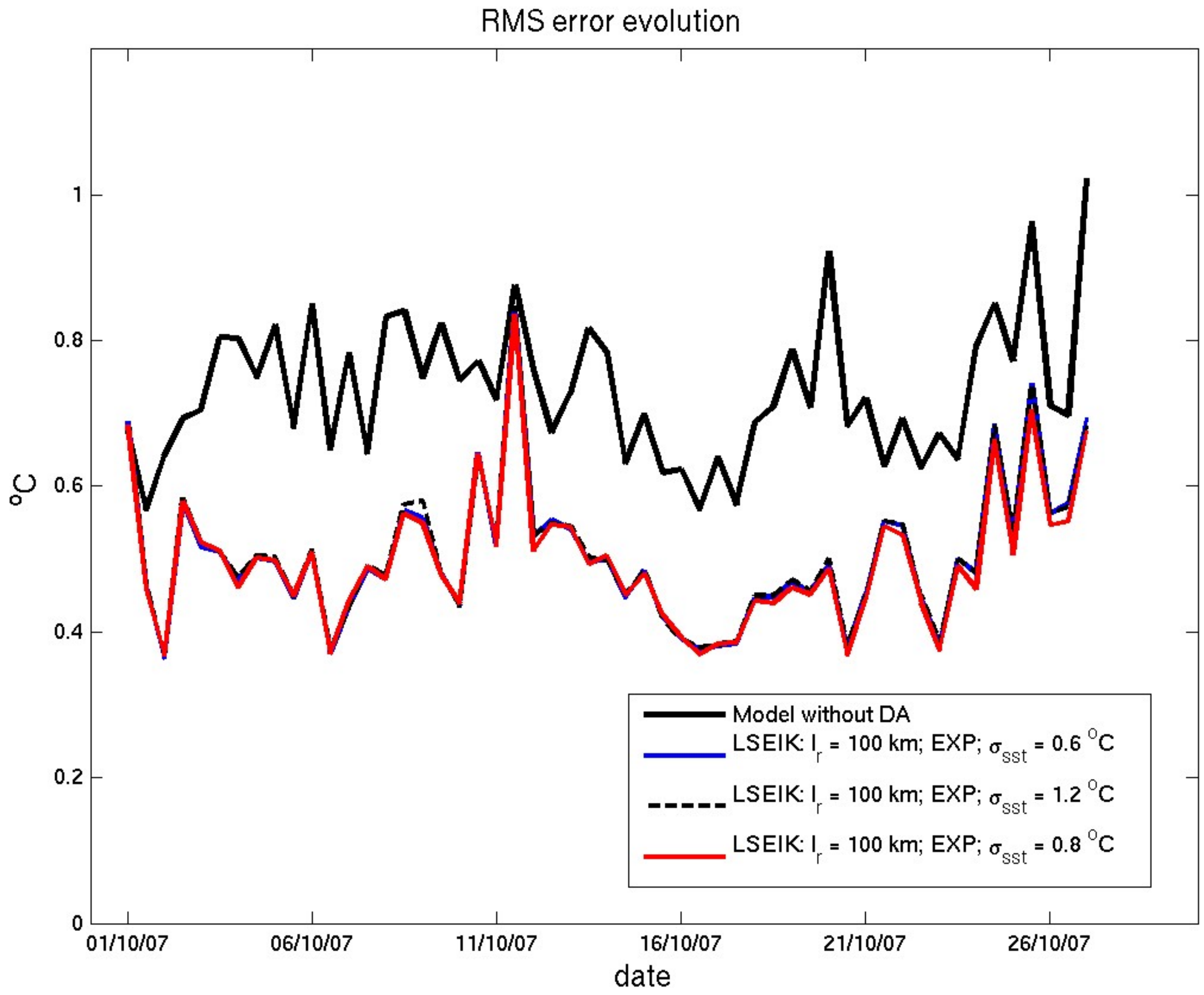
Use SVD decomposition

Improvement of SST analysis and forecast



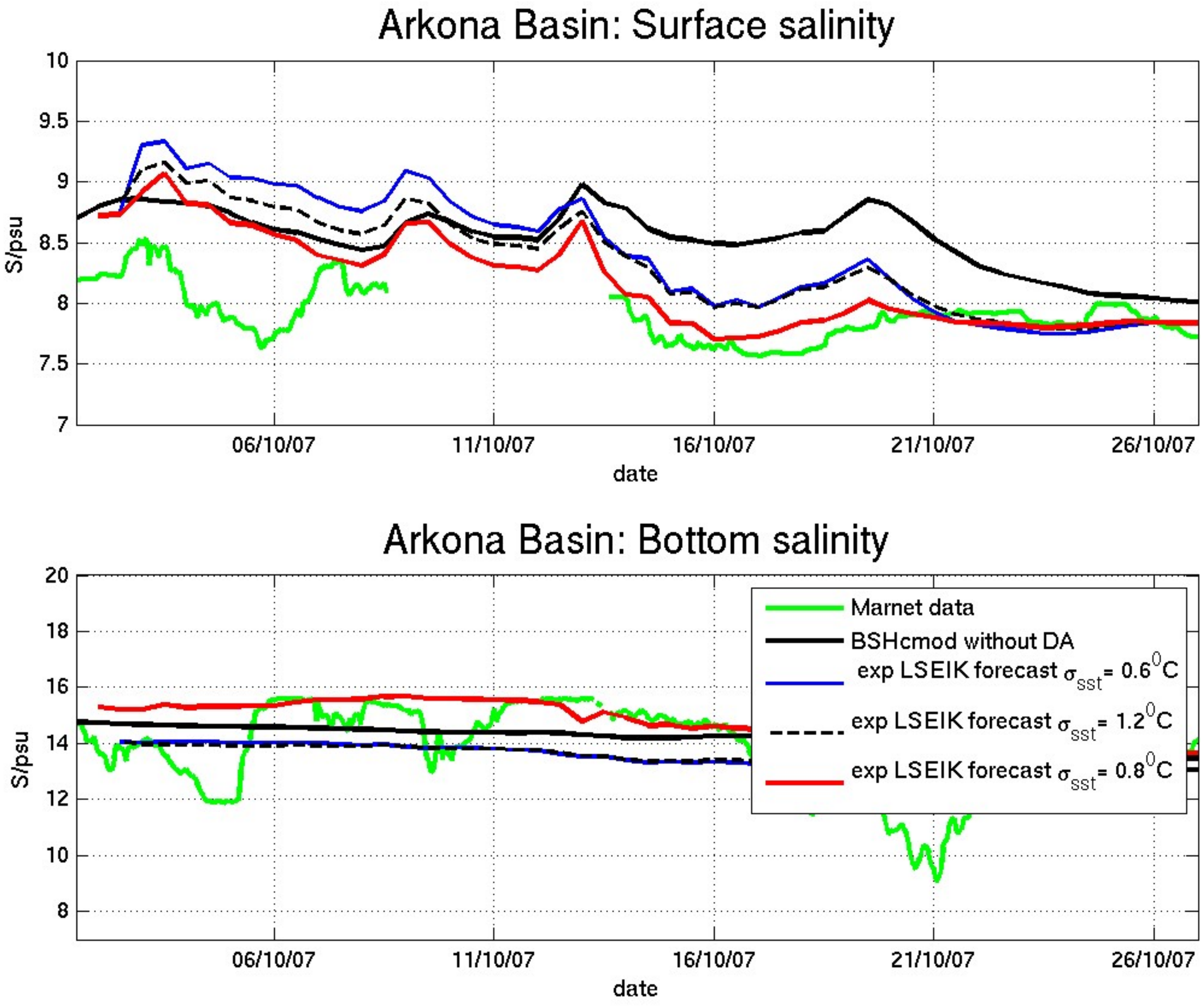
Improvement of SST forecast

Experiment	$He(M)$
$\sigma_{sst}=0.6\,^oC$	3.99
$\sigma_{sst}=0.8\,^oC$	4.33
$\sigma_{sst}=1.2\,^oC$	3.90



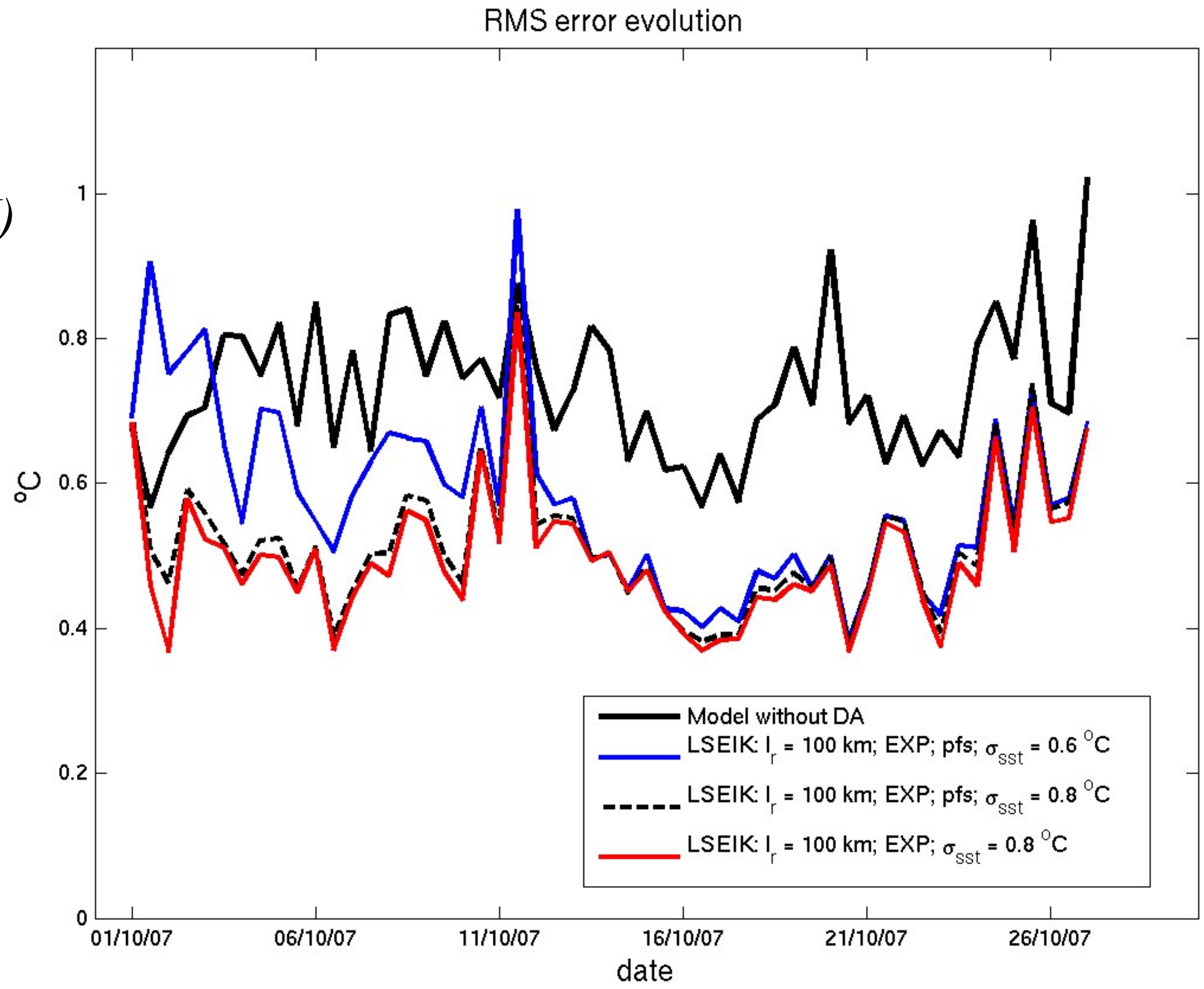
Comparison with independent information

Experiment	$He(M)$
$\sigma_{sst}=0.6\,^{o}C$	3.99
$\sigma_{sst}=0.8\,^{o}C$	4.33
$\sigma_{sst}=1.2\,^{o}C$	3.90



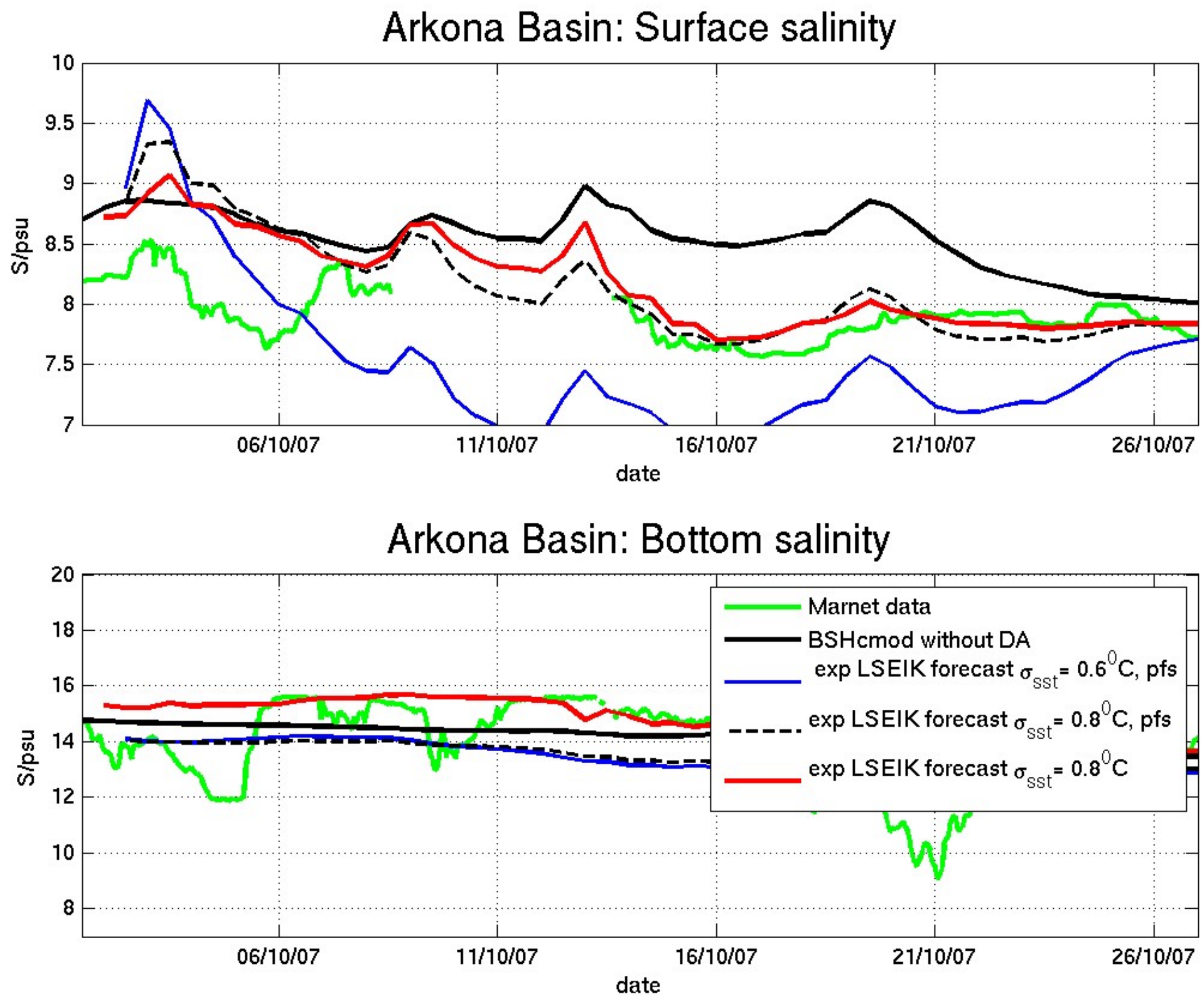
Sensitivity of the forecast quality

Experiment	$He(M)$
$\sigma_{sst}=0.6^oC, P^{fs}$	3.57
$\sigma_{sst}=0.8^oC, P^{fs}$	4.17
$\sigma_{sst}=0.8^oC$	4.33



Comparison with independent information

Experiment	$He(M)$
$\sigma_{sst}=0.6^oC,\ P^{fs}$	3.57
$\sigma_{sst}=0.8^oC,\ P^{fs}$	4.17
$\sigma_{sst}=0.8^oC$	4.33

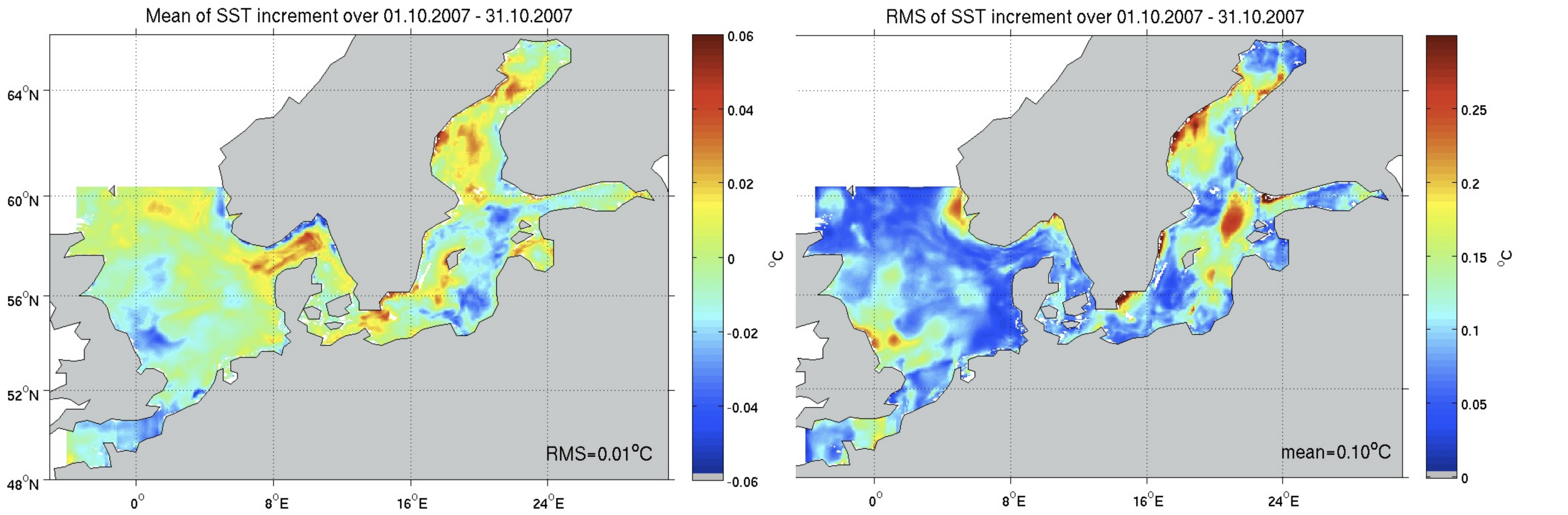


Comparison with independent information

Deviation from MARNET SST Daten

Station	RMS (°C)			Bias (°C)		
	Model	LSEIK	NOAA	Model	LSEIK	NOAA
Arkona	0.88	0.58	0.61	-0.29	0.	0.04
Darβ	1.27	0.81	0.69	-0.55	-0.17	0.01
Kiel	0.79	0.49	0.61	-0.13	0.07	0.08
Fehm	0.63	0.43	0.56	-0.16	0.03	0.16
Ems	0.67	0.45	0.49	0.33	0.2	0.17
Dbucht	0.97	0.53	0.57	-0.34	-0.03	0.27
nsb			0.73			

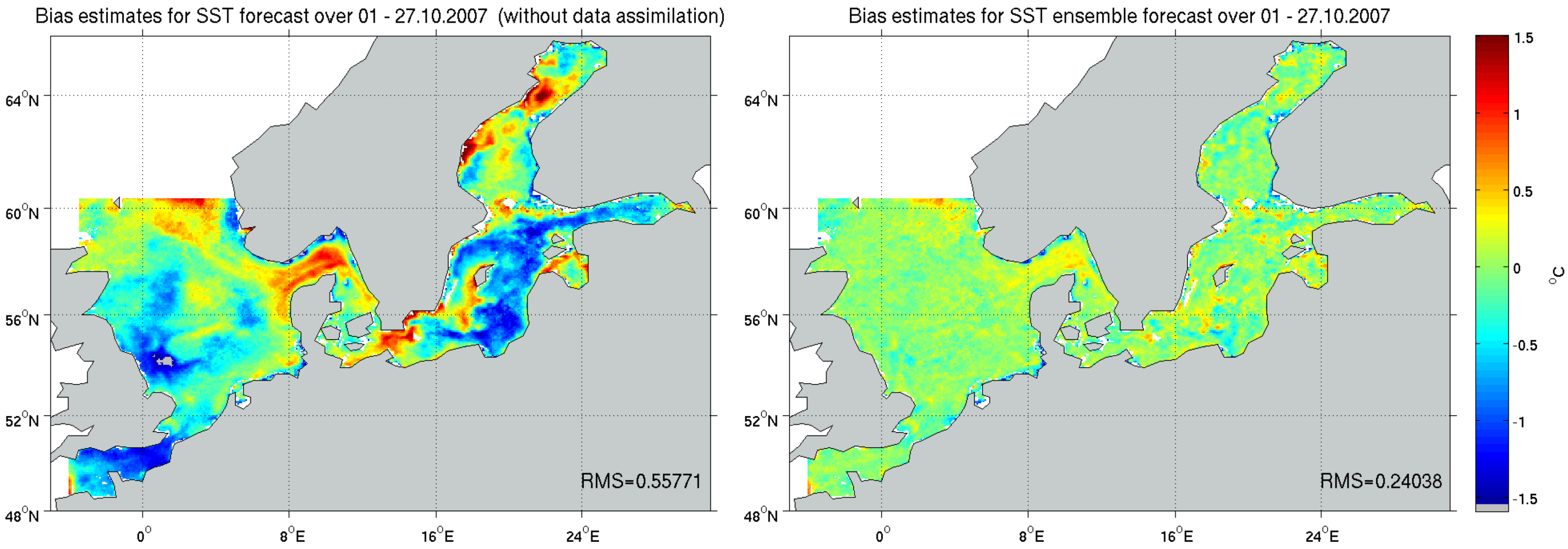
Increment Analysis



Improvement of SST forecast in the North and the Baltic Seas when sequentially assimilating satellite data

AWI

Bias without DA with LSEIK filter



Bias reduction

Conclusions

AWI

We have demonstrated two examples of the PME implementation for a posteriori estimating the model and data errors in data assimilation problem.

The chlorophyll satellite data assimilation based on a posteriori choosing of the data weight allowed us to compare the quality of the data and ecosystem model prediction and discern the low quality of the satellite data for high latitudes and for the coastal region of the North Atlantic.

The procedure of the secondary inversion of biogeochemical fluxes makes it possible to restore the mass balance broken while performing the weak constraint parameter estimation and to refine the estimates of the biogeochemical fluxes.

The spatial distribution of the biogeochemical parameters is in a good agreement with independent information about spices composition/distribution and their physiology.

Implementation of the PME for assessing prior model and data error statistics in SST data ensemble based assimilation for an operational forecasting model of the North and Baltic Seas revealed the best agreement of the forecast with independent data under the assumptions on initial model and data error statistics, which produced the ME of the posterior distribution.

Investigation of the PME implementation in a local analysis content is of our further interest.

References

Boltzmann, L., 1964: Lectures on Gas Theory. Cambridge University Press, 490 pp. [First published as Vorlesungen über Gastheorie, Barth, 1896.].

Gibbs, J. W., 1902: Elementary Principles in Statistical Mechanics. Yale University Press, 207 pp.

Kivman, G. A., Kurapov, A. L., Guessen, A. V., 2001: An Entropy Approach to Tuning Weights and Smoothing in the Generalized Inversion. *J. Atmos. Oceanic Technol*., 18, 266–276.

Losa, S. N, Kivman, G. A., Ryabchenko, V. A., 2004: Weak constraint parameter estimation for a simple ocean ecosystem model: what can we learn about the model and data?, *Journal of Marine Systems*, Volume 45, Issues 1-2, Pages 1-20, ISSN 0924-7963, 10.1016/j.jmarsys.2003.08.005.

Losa, S. N., Vézina, A., Wright, D., Lu, Y., Thompson, K., Dowd, M., 2006: 3D ecosystem modelling in the North Atlantic: Relative impacts of physical and biological parameterizations. *Journal of Marine Systems*, Volume 61, Issues 3-4, Pages 230-245, ISSN 0924-7963, 10.1016/j.jmarsys.2005.09.011.

Nerger, L., S. Danilov, W. Hiller, and J. Schröter. Using sea level data to constrain a finite-element primitive-equation model with a local SEIK filter. Ocean Dynamics 56 (2006) 634

Shannon, C. E., 1948: A mathematical theory of communication. *Bell Syst. Tech. J.,*27, 379–423, 623–655.

Tarantola, A., 1987: Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation. Elsevier, 613 pp.

van Leeuwen, P. J., and G. Evensen, 1996: Data assimilation and inverse methods in terms of a probabilistic formulation. *Mon. Wea. Rev.,*124, 2898–2913.