Oceanologia (2015) 57, 17—24

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/oceano

ORIGINAL RESEARCH ARTICLE

Assimilation of the satellite SST data in the 3D CEMBS model☆

Artur Nowicki [a,*], Lidia Dzierzbicka-Głowacka [a], Maciej Janecki [a], Maciej Kałas [b]

[a] *Institute of Oceanology, Polish Academy of Sciences, Sopot, Poland*
[b] *Marine Institute in Gdańsk, Gdańsk, Poland*

Received 4 July 2014; accepted 4 July 2014
Available online 22 October 2014

KEYWORDS
Satellite data assimilation;
Marine ecosystem modelling;
Baltic Sea;
Operational oceanography

Summary The 3D CEMBS (3D Coupled Ecosystem Model of the Baltic Sea) is a coupled ecosystem model of the Baltic Sea. In operational mode it computes 48-h forecasts of the hydrodynamic and biochemical parameters describing the Baltic Sea state. The Cressman assimilation scheme was implemented as part of the system in order to improve overall model accuracy. The system uses satellite-measured sea surface temperature from the MODIS Aqua spectroradiometer for the assimilation process. The satellite measured SST is obtained from a predefined server, which is part of the Satellite Monitoring of the Baltic Sea Environment project (SatBałtyk).

To validate the model results and the impact of assimilation on the model's accuracy, two separate test runs were performed using historical data covering the years 2011 and 2012. Independent computations were performed for the model with and without satellite SST assimilation, respectively referred to in this paper as 3D CEMBS_A and 3D CEMBS. The results of the computations were then compared with satellite and in situ measured data to validate the model and the assimilation scheme's implementation.

The objective of this paper is to describe the implementation of the satellite SST data assimilation algorithm and to present the results of the preliminary validation of the models with observations.

☆ The partial support for this study was also provided by the project Satellite Monitoring of the Baltic Sea Environment — SatBałtyk founded by European Union through European Regional Development Fund contract no. POIG 01.01.02-22-011/09.
* Corresponding author at: Institute of Oceanology, Polish Academy of Sciences, Powstańców Warszawy 55, 81-712 Sopot, Poland. Tel.: +48 58 7311915.
E-mail address: anowicki@iopan.gda.pl (A. Nowicki).
Peer review under the responsibility of Institute of Oceanology of the Polish Academy of Sciences.

ELSEVIER Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.oceano.2014.07.001

1. Introduction

Numerical modelling of the Baltic Sea basin is a complicated problem. Many factors have to be taken into account, such as the inflow of waters from the North Sea, as well as the influence of rivers and atmospheric conditions. The vertical parameterization must be very accurate as the distinct stratification of the Baltic Sea is very important. Atmospheric data must also be of the highest quality as they are the main forcing fields of the model. Even meeting all these requirements does not guarantee that the model itself will be able to produce good quality results, close to the real state, over a long period of time. This is why satellite data assimilation is a very important matter that needs to be implemented to constrain the model with observations. There are many different methods of satellite data assimilation used worldwide. The Cressman analysis scheme (Cressman, 1959) is one of the simplest but also one of the fastest methods, which is important, as the main aim of the 3D CEMBS (3D Coupled Ecosystem Model of the Baltic Sea) is to produce forecasts in operational mode. This was the main argument for choosing this method over other more complicated methods that require much more computing power and time. Following its validation, the assimilation procedure was implemented into the operational mode of the model. This version of the model provides data for the operational system of the SatBałtyk project, described in detail by Woźniak et al. (2011a,b). The results of this work together with the operational system's configuration are presented in this paper.

2. Material and methods

2.1. Assimilation scheme

Data assimilation is an analysis that combines time-distributed observations and a dynamic model. This kind of analysis gives much better results than simpler methods like the spatial interpolation of observations. According to the way in which the updating is done in time, data assimilation can be divided into variational and sequential data assimilation. In the first approach, past observations until the present time are used simultaneously to correct the initial conditions of the model. In sequential assimilation, observed data are used as soon as they appear in order to correct the model state. There are many different methods of introducing the observed data into the model, from the Cressman scheme, through Optimal Interpolation, 3-D and 4-D variational methods, to different modifications of the Kalman Filter. As the 3D CEMBS operational system uses the Cressman scheme, other methods will not be presented in greater detail in this paper.

The Cressman method is a simple and computationally fast assimilation scheme, which makes it a good choice for a data assimilation system used to create forecasts in operational mode. It is also very accurate in comparison to its low complexity. Its main disadvantage is that it may produce unrealistic extrema in the grid values near the edges of the spatial domain. It can be also unstable if the model grid density is higher than the observation grid density. However, in the case of satellite data this is not an issue, as the spatial resolution of the satellite data used is higher than the model grid resolution.

The Cressman method comes down to few simple steps that are performed as follows. Firstly, the background state x_b is set equal to the previous forecast performed by the model. Then the satellite data used for the assimilation are stored in the matrix denoted by y. Data suspected of being invalid because of clouds, the presence of ice or any other reason are masked out. The result of the analysis x_a is then calculated according to the following equation:

$$x_a(j) = x_b(j) + \frac{\sum_{i=1}^{n} w(i,j)\{y(i) - x_b(i)\}}{\sum_{i=1}^{n} w(i,j) + E^2},$$

where i and j represent the satellite and model data grid-points respectively, and $d_{i,j}$ is the distance between points i and j. The main parameters of the Cressman method that need to be chosen are the influence radius R and the shape of the weight function w, which determine how the satellite data influence the model. One of the disadvantages of this method is that the influence radius has to be determined by trial and error; this makes parameterization of this method laborious. After many trials with different sets of the parameters, the one that gave the best results was chosen. The radius R of the influence was set to 20 grid-points. Beyond that distance the satellite data weight equals zero. The weight function in this case is equal to:

$$w(i,j) = \max\left(0, \frac{R^2 - d_{i,j}^2}{R^2 + d_{i,j}^2}\right).$$

In addition, the parameter E used in the successive correction method was introduced. E^2 is an estimate of the ratio of the observation error to the first guess field error. E was set to 0.5 ($E^2 = 0.25$), which means that the satellite data are treated as more accurate than the model data. However, they never have a weight equal to one. In the absence of this parameter ($E^2 = 0$), the satellite data, if present at a particular location, would be given a weight of one. This means that the model data at this point would be omitted. The presence of E^2 ensures that the model data are taken into account everywhere and ensures smoothing of the analysis product, which prevents possible instabilities. The product of assimilation is then used as the new initial state of the model from which the new forecast is calculated.

2.2. The operational model system

The current version of the 3D CEMBS (3D Coupled Ecosystem Model of the Baltic Sea) is based on the CESM (Community Earth System Model) developed at the National Center for Atmospheric Research. It was adapted for the Baltic Sea region as a coupled sea-ice model consisting of POP (The Parallel Ocean Program) and CICE (The Los Alamos Sea Ice Model). Atmospheric fields from the ICM (Interdisciplinary Centre for Mathematical and Computational Modelling) of Warsaw University are used to force the model together with historical data of river inflows. 71 main rivers are taken into account. All these components are coupled by a CPL7 (Coupler, version 7), which controls time and data exchange between these components. The model is configured in a horizontal resolution of 1/48 degrees and it is divided into 21 vertical levels. In the first half of 2013 the Cressman analysis scheme was used to implement satellite SST data assimilation. The data gathered in the

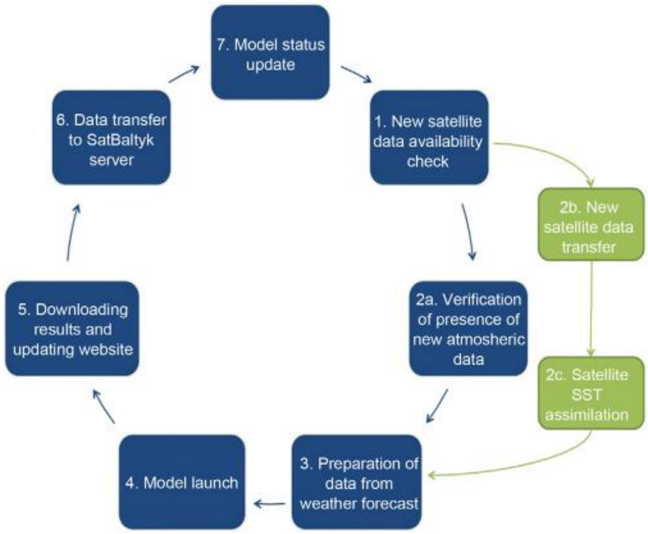

Figure 1 3D CEMBS_A with operational assimilation working scheme.

SatBałtyk project were used as the source of satellite data. The aim of this implementation was to improve the model's accuracy. The model and satellite data are complementary to each other as in the case of high cloud coverage over the Baltic Sea the model is the main source of data. The 3D CEMBS_A model is currently running in operational mode. This mode is split into two separate sub-modes. The regular mode produces 48-h forecasts using new weather forecasts from the ICM as forcing fields. The forecasts are produced on a regular basis every 6 h. The hydrodynamic part of the model produces sea temperature, salinity, current speed and direction, sea surface height, ice area cover and ice thickness (Dzierzbicka-Głowacka et al., 2013a). It also provides several biological, chemical and ecological parameters (Dzierzbicka-Głowacka et al., 2013b). Results are then stored in the local archive and posted to a model website. The parameters from the surface are interpolated to 1 km resolution, uploaded onto the SatBałtyk server and are available from the project's website. The second mode is the assimilation mode. This remains idle until a new set of satellite data is available on the SatBałtyk server, at which time the system switches to assimilation mode. It performs data assimilation, sets the assimilated data as the new initial state of the model and performs new calculations from the time of the satellite data's appearance until the current forecast ending time. Afterwards the system uploads new results in the same way as in the regular mode. Then it switches back to regular mode. The Fig. 1 outlines the scheme of how the system operates.

3. Results and discussion

The test run of the model was performed on the historical data covering the years 2011 and 2012. Independent calculations were performed for the model with and without satellite SST assimilation, respectively referred to in this paper as 3D CEMBS_A and 3D CEMBS. The results of both runs were compared with each other as well as with satellite data and different in situ measurements. Validation of the satellite data assimilation with the 3D CEMBS model consisted of two parts. Firstly, the results of both models were compared with the satellite data to check whether the assimilation algorithm was working properly and to examine the impact of the assimilation on the model results. Then, the results from both

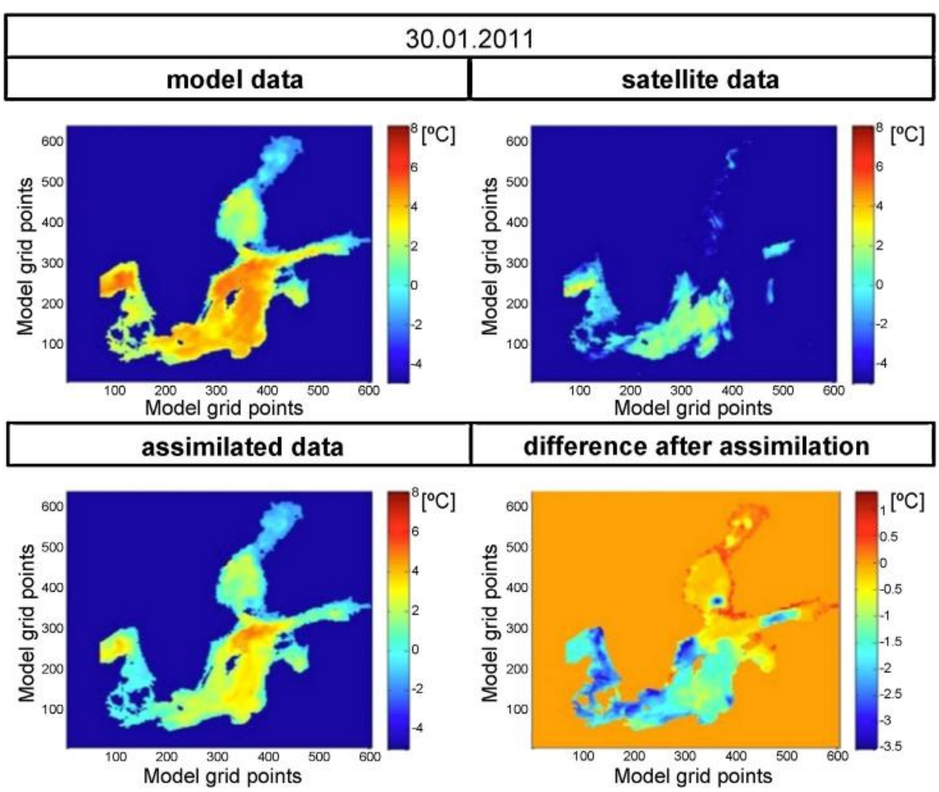

Figure 2 Comparison of sample results from models with and without assimilation of satellite data.

model test runs were compared with different in situ data to check whether the assimilation actually improved the overall model accuracy. For a preliminary assessment of the correctness of the assimilation algorithm, sample images from the satellite were compared with the results of both models from different days. Fig. 2 shows the sample scene from January 1st, 2011. The figure consists of the model data before assimilation, the satellite data used for assimilation and the model data after satellite data assimilation. The picture at bottom right shows the difference between the two models. In this example the satellite measured temperature is mostly lower than the one calculated by the model before assimilation. Assimilation lowers the temperature in the model surface layer, as expected. The same results were obtained for other scenes, which indicates that the assimilation algorithm is working properly. Of course, visual comparison is not sufficient, so additional tests were performed. In order to assess the accuracy of the assimilation algorithm and model accuracy, statistical parameters such as the correlation coefficient r, the mean systematic error $\langle\varepsilon\rangle$ and the standard deviation $\langle\sigma\rangle$ between both models and satellite data were calculated for all data from the years 2011 and 2012, as were the mean values and differences between the models. After validation of the assimilation algorithm, the same methods were used to assess the model error with respect to in situ data. The parameters were calculated according to the principles of arithmetic statistics presented by following equations:

- absolute mean error (systematic):

$$\langle\varepsilon\rangle = \frac{\sum_{i=1}^{n}(X_i - Y_i)}{N},$$

- standard deviation (statistical error):

$$\sigma_\varepsilon = \sqrt{\frac{\sum_{i=1}^{n}\left(\varepsilon_i - \langle\varepsilon\rangle\right)^2}{N}}.$$

In the above equations X_i and Y_i are model and in situ data, $e_i = X_i - Y_i$. The results of the assessment are presented below. Table 1 lists the calculated statistical parameters. The statistics provide clear confirmation that the assimilation algorithm does improve accuracy. Correlation of the model results with the assimilation and the satellite data is better and the errors are smaller. Figs. 3—7 illustrate calculated mean values and differences of surface layer temperature from both models. Fig. 3 shows the mean value over time through the years 2011 and 2012. The black points represent the 3D CEMBS model and the grey ones the 3D CEMBS_A model. One can see that the surface layer responds more slowly to the weather conditions in the model without assimilation; hence, the temperature of this layer is lower in spring and summer and slightly higher in autumn and winter

Table 1 Statistical comparison of both models with satellite measured SST.

Model type	R	$\langle\varepsilon\rangle$ [°C]	$\langle\sigma\rangle$ [°C]
3D CEMBS	0.949	−1.36	2.01
3D CEMBS_A	0.976	−0.72	1.35

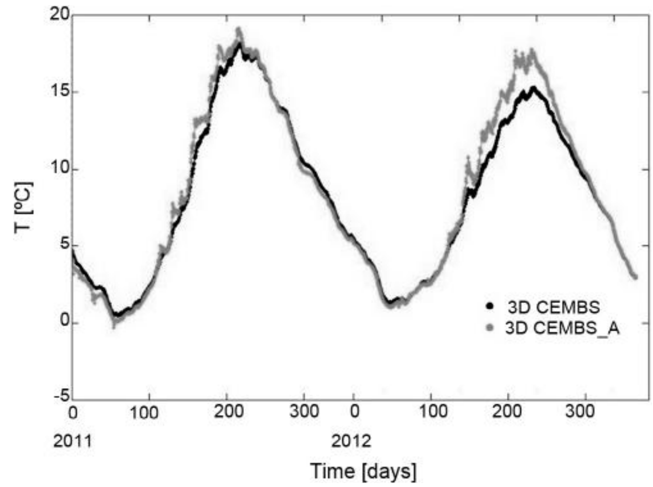

Figure 3 Mean surface layer temperature over the model domain in the years 2011—2012.

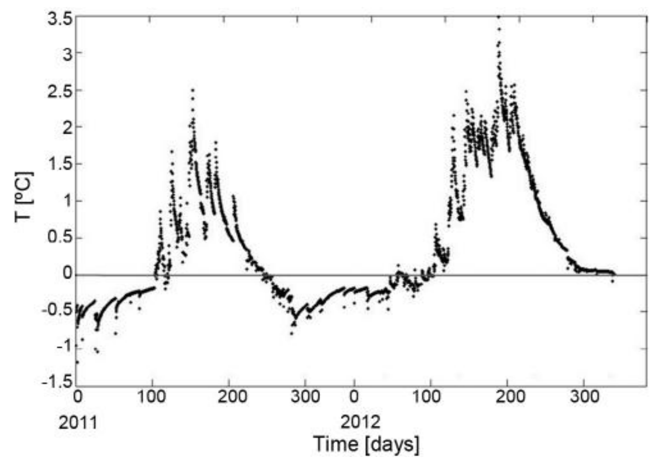

Figure 4 Mean surface layer temperature difference over the model domain in the years 2011—2012.

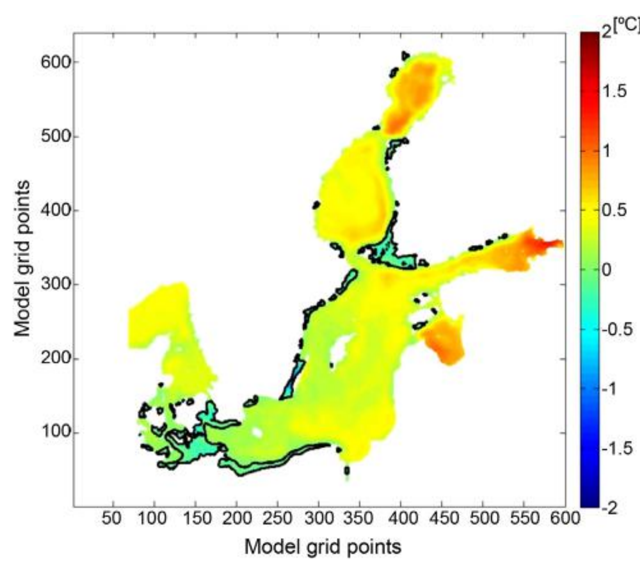

Figure 5 Mean difference between the surface layer temperature of the 3D CEMBS and 3D CEMBS_A models over the years 2011—2012.

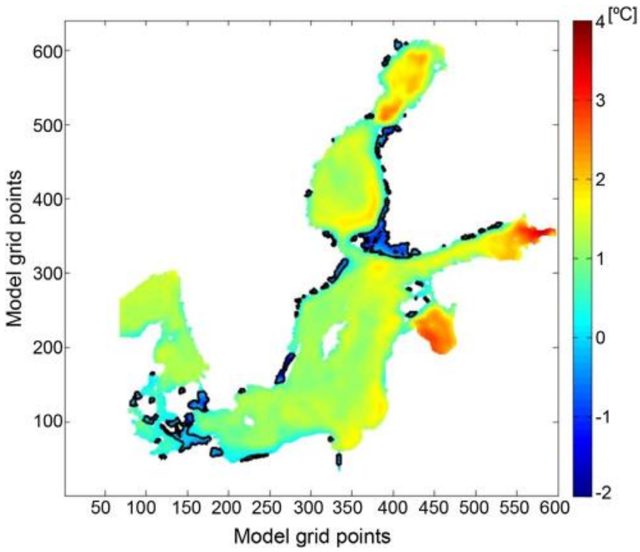

Figure 6 Mean difference between the surface layer temperature of the 3D CEMBS and 3D CEMBS_A models during warm seasons.

than the temperature computed using assimilation. To distinguish these two periods better, Fig. 4 shows the difference between both model temperatures. Positive values mean that the model with the assimilation of satellite measured SST gives higher temperature values. This difference is due to the definition of the surface layer in the model, namely, the first layer of the model grid. The thickness of this layer is 5 m, which is much more than that of the actual surface layer for which the SST is measured from satellite instruments. Of course, a thicker layer responds more slowly to atmospheric forcing, as it has a higher heat capacity. Assimilation of the satellite data into the model adjusts the surface layer temperature to the atmospheric conditions faster; hence, it responds faster to the changes. Fig. 4 shows the temporal differences in temperature between the two models. On the other hand, Fig. 5 shows the spatial distribution of changes in the surface temperature driven by the satellite data assimilation. The figure presents the mean temperature difference between the two models. The black line divides areas with positive and negative temperature changes. Positive values mean that assimilation caused an increase of the model temperature, while negative values mean a decrease of temperature as a result of assimilation. Most of the model domain is covered with positive values. This means that, in general, assimilation of satellite SST causes a rise in the model temperature. In other words, the model bias is mostly negative. To investigate the impact of assimilation on the results more precisely, the tested data were divided into two periods, based on Fig. 4. These periods will later be referred to as the warm and cold seasons. The results are presented in Figs. 6 and 7. Assimilation also increased the correlation of the data, which is reflected in Table 1 and Fig. 8. Figs. 6 and 7 show that during the warmer season of the year the model bias is much greater than during the cold season, when it is close to zero. This is also confirmed in Fig. 4. This may be due partially to the fact that there is much less information from the satellite during winter. Further information emerging from Figs. 6 and 7 is that there are regions where the model results are always lower than the satellite measurements. These are the regions with positive values on both figures such as the Gulf of Finland and parts of the Gulf of Bothnia. One can also see that the differences between the two models are much lower near seashores. This is probably due to the lower depth in coastal zones, as the temperature of shallower water can change faster. Fig. 9 presents a correlation of model data from both cases with satellite measured SST. The points selected for the comparison were based on the locations of data from the ICES database shown in Fig. 8.

The influence of the SST assimilation on the other parameters such as salinity, currents and sea surface height was also investigated. However, these parameters did not show any

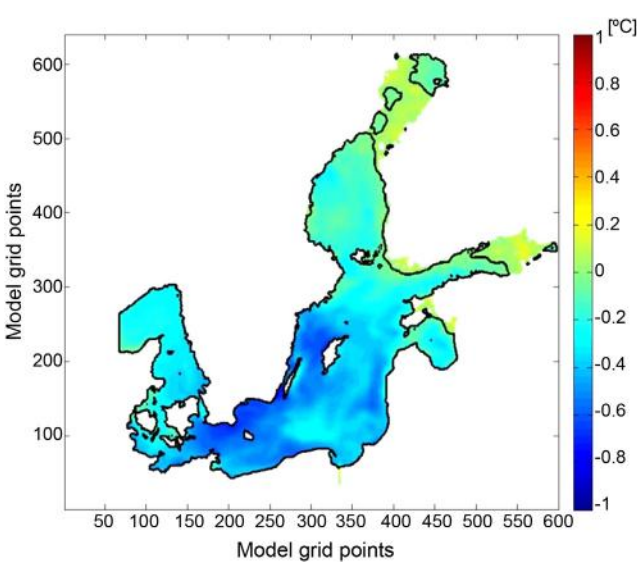

Figure 7 Mean difference between the surface layer temperature of the 3D CEMBS and 3D CEMBS_A models during cold seasons.

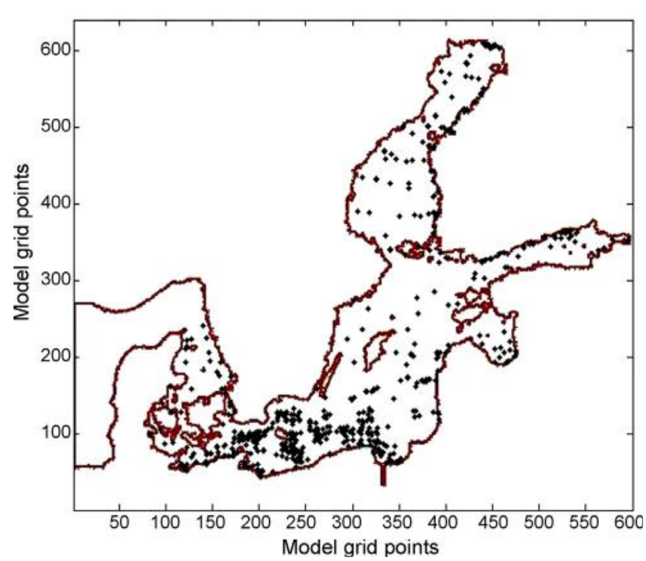

Figure 8 Map of locations of in situ data from the ICES database.

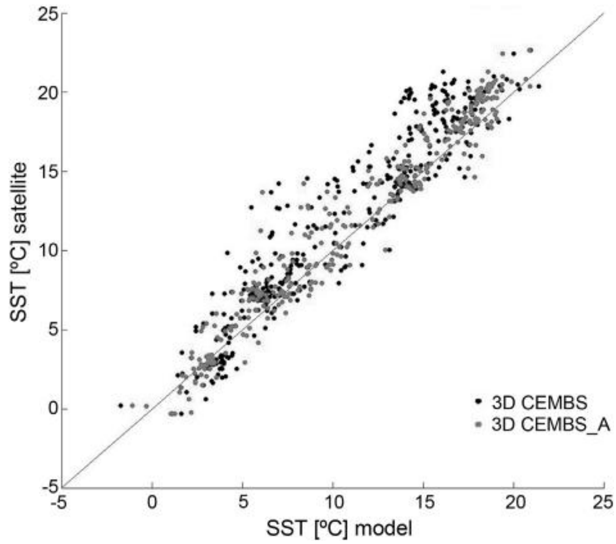

Figure 9 Correlation of the sea surface temperature from the 3D CEMBS and 3D CEMBS_A models with satellite measurements.

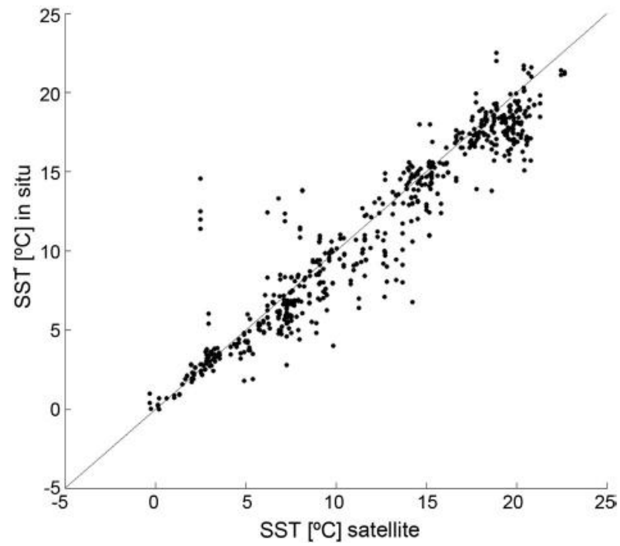

Figure 10 Correlation of the satellite measured SSTwith in situ data.

meaningful differences. Even during the period with the greatest differences between 3D CEMBS and 3D CEMBS_A in the computed temperature, that is, in summer 2012, the other parameters varied only slightly.

After positive validation of the assimilation algorithm's performance, both model results could be compared with a set of in situ data to estimate the actual influence of the assimilation. The in situ data used for the comparison were obtained from the ICES database. This part of the validation also covered data from different locations in all parts of the Baltic Sea from 2011 to 2012. The locations of the in situ data are marked in Fig. 8. Table 2 presents the results of the statistical analysis of the data. The not-assimilated model results have a negative bias with respect to the in situ data, but it is significantly smaller in comparison to results from Table 1. This means that the satellite measurements give a higher temperature than that measured in situ. This is confirmed by the positive bias of the satellite data with respect to the in situ measurements. Nevertheless, assimilation of the satellite measured SST improves the accuracy of the model, which is confirmed by the results presented in the last row of Table 2. Figs. 10 and 11 present a correlation of the in situ results with the results from remote sensing and both versions of the model.

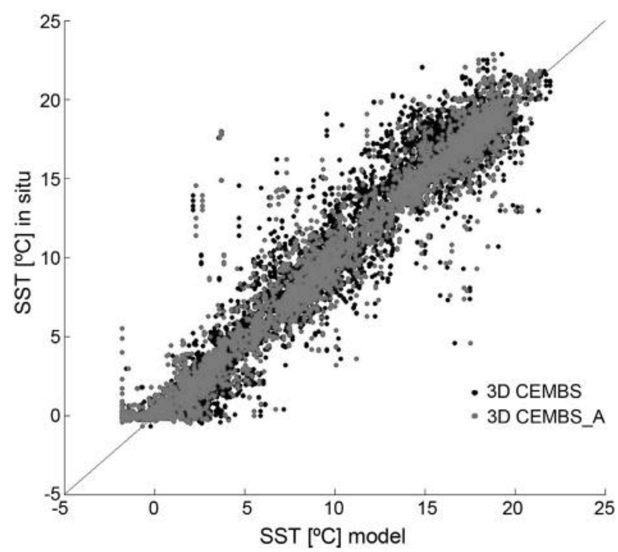

Figure 11 Correlation of the surface temperature from the 3D CEMBS and 3D CEMBS_A models with in situ measurements.

The statistics show the average performance of the assimilation algorithm over the whole year. This means that the data are dominated by the main seasonal signal. Removal of this signal from the data reveals the model's accuracy in greater detail. Table 3 lists the statistics of both models after removal of the seasonal signal. This shows clearly that assimilation of the satellite measured SST has a positive impact on the model simulations. The correlation coefficient, when

Table 2 Statistical analysis of both models and satellite data with in situ measured SST.

Compared data	r	$\langle\varepsilon\rangle$ [°C]	$\langle\sigma\rangle$ [°C]
Satellite vs. in situ	0.958	0.59	1.77
3D CEMBS vs. in situ	0.957	−0.21	1.90
3D CEMBS_A vs. in situ	0.973	−0.06	1.53

Table 3 Statistics of the model results with and without assimilation in comparison with in situ measured SST after removal of the main seasonal signal.

Compared data	r	$\langle\varepsilon\rangle$ [°C]	$\langle\sigma\rangle$ [°C]
3D CEMBS vs. in situ	0.64	0.24	1.81
3D CEMBS_A vs. in situ	0.73	0.09	1.59

Table 4 Statistics of the models with and without SST assimilation in comparison with in situ measurements for each month separately.

Month	Measurements	3D CEMBS		3D CEMBS_A	
		$\langle\varepsilon\rangle$ [°C]	$\langle\sigma\rangle$ [°C]	$\langle\varepsilon\rangle$ [°C]	$\langle\sigma\rangle$ [°C]
1	297	1.62	1.44	1.02	1.07
2	531	0.17	1.69	−0.30	0.85
3	688	0.21	1.16	−0.09	0.17
4	366	−0.49	2.22	−0.66	0.80
5	848	−0.46	1.92	−0.10	1.80
6	544	−1.44	1.83	−0.67	2.02
7	514	−1.26	2.06	−0.39	1.75
8	1010	−0.62	2.05	0.05	0.18
9	489	0.02	1.45	0.11	1.07
10	394	0.68	1.64	0.24	0.86
11	402	0.44	1.27	0.32	0.10
12	175	0.84	1.09	0.61	1.53

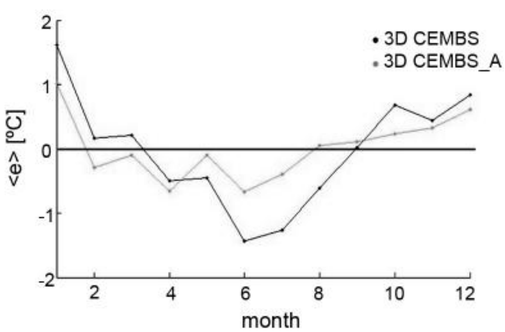

Figure 12 Mean monthly systematic error of the sea surface temperature calculated from the 3D CEMBS and 3D CEMBS_A models.

not dominated by the seasonal signal, changes significantly more after assimilation is implemented. The systematic and statistical errors are similar to those prior to the removal of this signal.

To provide more detailed results showing the performance of the models in different months of the year, the main statistical parameters were calculated for each month separately. This gives a better insight into the model and the

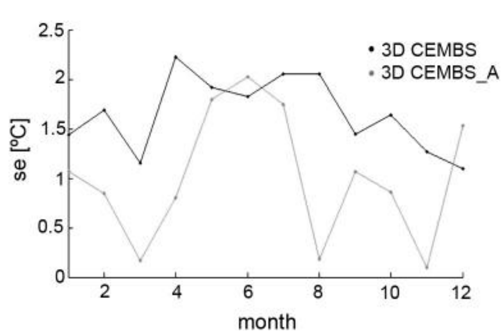

Figure 13 Mean monthly standard deviation of the sea surface temperature calculated from the 3D CEMBS and 3D CEMBS_A models.

assimilation results in different seasons. Figs. 12 and 13 and Table 4 give the results of these calculations. As one can see, the systematic error after assimilation is closer to zero, which confirms previous findings about the effectiveness of the assimilation algorithm. The shape of the plot indicates that during colder seasons of the year the model is positively biased and that during spring and summer its bias is negative. This corroborates earlier conclusions that the surface temperature calculated by the model changes more slowly than the one measured in situ or by satellite. The standard deviation does not show any clear dependence on time of year; nevertheless, SST assimilation decreased its value in most months.

4. Conclusion

Application of the Cressman assimilation algorithm into the 3D CEMBS_A model improved its accuracy and conformance of its results with in situ and satellite measured SST. Analysis of the results gives a better view of the spatial and temporal error distribution in the investigated period of time. Overall, the statistics show an increase in model correlation with the satellite data from ca 0.95 for the 3D CMEBS model to ca 0.98 for 3D CMEBS_A. Also, the mean arithmetic error and standard deviation are smaller for the model with SST assimilation, which confirms the assimilation algorithm's correctness. Similar results are obtained when the models are compared with in situ data. The correlation coefficient in this case increased from 0.957 to 0.973 and the systematic error decreased strongly in value. In addition, the standard deviation decreased in value slightly. After removal of the main seasonal signal, the statistics of the model results presented in Table 3 reveal an even bigger difference in correlation between the two models and the in situ data. The simulations of SST are also better with respect to monthly means, as shown in Table 4 and Figs. 12 and 13. Assimilation of satellite data into the 3D CEMBS_A model is therefore reasonable, as is its further development. The ongoing development of the SST assimilation system as well as other parameters such as chlorophyll a is included in our research plans.

Acknowledgements

The computing presented in this paper was carried out on the Galera super computer at the Academic Computer Centre in Gdansk (CI TASK).

In situ data used for validation were obtained from ICES Dataset on Ocean Hydrography. The International Council for the Exploration of the Sea Copenhagen. 2011, http://ocean.ices.dk/helcom/Helcom.aspx?Mode=1.

References

Cressman, G.P., 1959. An operational objective analysis system. Mon. Weather Rev. 87, 367—374.

Dzierzbicka-Głowacka, L., Jakacki, J., Janecki, M., Nowicki, A., 2013a. Activation of the operational ecohydrodynamic model (3D CEMBS) — the hydrodynamic part. Oceanologia 55 (3), 519—541.

Dzierzbicka-Głowacka, L., Jakacki, J., Janecki, M., Nowicki, A., 2013b. Activation of the operational ecohydrodynamic model (3D CEMBS) — the ecosystem module. Oceanologia 55 (3), 543—572.

Woźniak, B., Bradtke, K., Darecki, M., Dera, J., Dudzińska-Nowak, J., Dzierzbicka-Głowacka, L., Ficek, D., Furmańczyk, K., Kowalewski, M., Krężel, A., Majchrowski, R., Ostrowska, M., Paszkuta, M., Stoń-Egiert, J., Stramska, M., Zapadka, T., 2011a. SatBałtyk — a Baltic environmental satellite remote sensing system — an ongoing project in Poland. Part 1: Assumptions, scope and operating range. Oceanologia 53 (4), 897—924.

Woźniak, B., Bradtke, K., Darecki, M., Dera, J., Dudzińska-Nowak, J., Dzierzbicka-Głowacka, L., Ficek, D., Furmańczyk, K., Kowalewski, M., Krężel, A., Majchrowski, R., Ostrowska, M., Paszkuta, M., Stoń-Egiert, J., Stramska, M., Zapadka, T., 2011b. SatBałtyk — a Baltic environmental satellite remote sensing system — an ongoing project in Poland. Part 2: Practical applicability and preliminary results. Oceanologia 53 (4), 925—958.