Assessment of sea surface temperature observational networks in the Baltic Sea and North Sea

Jun She⁎, Jacob L. Høyer, Jesper Larsen

DMI Centre for Ocean and Ice, Denmark

Received 23 September 2004; accepted 20 January 2005
Available online 14 November 2006

Abstract

The satellite and in situ Sea Surface Temperature (SST) observational networks in the Baltic Sea and North Sea are evaluated based on the quality of the gridded SST products generated from the networks. A multi-indicator approach is applied in the assessment. It includes evaluation of data quality, effective data coverage, field reconstruction error and model nowcast error. The results show that the best available full-coverage SST product is generated by assimilating the SST observations to obtain a yearly mean model bias of 0.07 °C and RMSE of 0.64 °C. The effective data coverage rate is 31% by using AVHRR (Advanced Very High Resolution Radiometer) data from NOAA (National Ocean and Atmosphere Administration) satellites 12, 14 and 16. The data redundancy increases rapidly with the number of infrared sensors. Using either NOAA satellite 12 or all 3 satellites makes a small difference with regard to derived effective coverage and the ocean model nowcast error. The influence of using the in situ SST observations in the SST field reconstruction is negligibly small. Instead, the major role of in situ SST observations is in calibrating the satellite observations. To study the relative importance of data quality and data coverage, an assessment is done for two satellite products: one product is based entirely on NOAA 12 data and has larger coverage but lower quality. The other product is a subset of the SAF products (derived from NOAA 14 and 16) and has lower coverage but higher quality. Based on current monitoring, modelling and assimilation technology, the results suggest that the data quality is an important factor in further improving the quality of the gridded SST products. Recommendations are made for possible further improvements of the existing SST observational networks.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Sea surface temperature; Observing system; Optimal design

1. Introduction

The sampling scheme and the monitoring technology are the two major components influencing the quality of a marine observational network. The accuracy of the measurements is dominated by the quality of the technological and operational implementation, while the total information content of the observational network depends mainly on its sampling scheme. A number of quantitative methods with different indicators have been used in evaluating the quality of sampling schemes and/or observational networks. These include: effective data coverage (She, 1996), noise–signal ratio (Smith and Meyers, 1996), sampling error (North and Nakamoto, She and Nakamoto, 1996a,b), field reconstruction error based on objective analysis (e.g., optimal interpolation method, She, 1996) and model sensitivity studies using sampled data, i.e., Observing System...
Experiment (OSE, e.g., *Atlas, 1997*) and Observing System Simulation Experiment (OSSE, e.g., *Kelly, 1997*).

For a general purpose of observational network assessment, the information product of a regional observational network is often a kind of gridded field (from one to four dimensions), e.g., a time series of SST fields in the Baltic–North Sea. The quality of the information product (i.e., the field reconstruction error) therefore reflects the quality of the sampling scheme and the observational network. However, the field reconstruction error is method-dependent. With different interpolation or modelling and assimilation methods, the field reconstruction error will be different. Conversely, indicators such as the effective data coverage, sampling error and noise–signal ratio are in principle method-independent.

The effective data coverage method estimates the size of a given area which is effectively covered with a given sampling scheme. This is normally based on a local characteristic scale analysis. Since this method was only documented in an internal technical report (*She, 1996*), details of the method will be described in the Section 5. The effective coverage method identifies the gaps and the effectively covered areas by an observational network, but it does not assess the quality of the information product derived from the measurements. Therefore, highly informative areas may not be resolved by the method. Furthermore, the measurement error is not included in the effective data coverage.

The Noise–signal ratio identifies the percentage of the signal explained by an observational network and its ratio to the noise level shown in the natural variability (*Smith and Meyers, 1996*). On the other hand, the sampling error method gives an estimate of the ratio of the local sampling uncertainty of the observational network to the local variance (*She and Nakamoto, 1996a,b*). Both methods are mainly based on characteristic scale analysis and spectrum analysis.

The statistical field reconstruction method reconstructs the information product by using objective analysis. It generates a field distribution of the reconstruction error, which identifies where more or less observations are needed. The statistical method can be applied efficiently to many different sampling schemes. However, these methods are based on linear statistics. The capability of model physics in removing noise and extracting non-linear correlations are not included.

Modelling has been an active part of atmosphere observing system assessment and designs through OSEs (Observation System Experiments) and OSSEs (Observation System Simulation Experiments). A comprehensive review can be found in *Atlas (1997)*. An OSE is an impact study carried out with existing observations. Normally two parallel model runs are carried out, one with and one without assimilating observations from the observing system. The resulting analyses and subsequent forecasts are then compared. An OSSE is similar to the OSE except that the ‘observations’ to be assimilated are simulated rather than real. The simulated observations are produced from a numerical model integration assumed to be the ‘known truth’. The OSEs have been used in numerical weather prediction for a couple of decades to quantify the usefulness of atmosphere observing systems (e.g. FGGE—First GARP Global Experiment observing system, *Bengtsson, 1981*). The advantage of the OSEs/OSSEs is that model dynamics is used in reconstructing the ocean status together with observations, which reduces the data requirements from observing systems. However, these experiments are very time consuming and costly. Furthermore, the OSE/OSSE methods are model and assimilation method dependent, which means that one may get different assessment results by using different models and assimilation methods for a given sampling scheme.

Based on the above discussions, this paper uses a multi-indicator method in order to make a comprehensive evaluation of the existing satellite and in situ SST observational networks in the Baltic–North Sea. The focus will be on the following issues: 1) the overall performance of the existing SST observational networks; 2) the relative importance of the different observational networks, with 1, 2, or 3 satellites and with or without in situ observations; and 3) the impacts of the network quality factors such as data coverage, data quality and reconstruction methods, on the assessment results.

The rest of the paper is organised as follows. Section 2 briefly describes the data quality and availability of the existing SST observational networks. Section 3 presents the SST natural variability and characteristic scales. Section 4 gives the ad hoc designed SST observational networks for the assessment. Results from the effective data coverage analysis, the statistical field reconstruction and the OSEs are shown in Sections 5–7. Finally, discussions and conclusions are found in Section 8.

### 2. Existing observing systems: data quality and availability

In this section, a description of existing Baltic–North Sea in situ and satellite observational networks, including the availability and quality of the measurements and pre-processing procedures, will be given in Sections 2.1 and 2.2. Section 2.3 describes the comparison between
satellite and in situ SST data based on a ‘bulk SST’ concept, including the results and impacts from sampling errors of the satellite and in situ SST observations.

2.1. In situ observational network

2.1.1. Spatial distribution

The in situ SST observations, used for validating the observations and correcting the skin SST to bulk SST, have been obtained from many different national monitoring agencies. The observations are measured by using Voluntary Observing Ships (VOS) linked to the Global Telecommunication System (GTS), moored buoys, Ferrybox, CTD casts, undulating profilers (e.g., Delphin) and thermalsalinographs (TSG). A check for redundant data was performed on all the available in situ observations to eliminate that the same observations were used twice. About 5% of the observations were redundant, using this method.

Fig. 1 shows all the in situ SST stations in 2001, which were collected in the EU Fifth Framework Program Project ODON (Optimal Design of Observational Networks). The GTS network is under the frame of WMO’s World Weather Watch (WWW). In the Baltic–North Sea, the GTS surface marine observations contain two major components: VOS and fixed platforms (e.g. moored buoys, lighthouses and oil platforms). However, not all the moored buoys are included in the GTS system. The CTD casts, underway SST measurements from TSG and the undulating profiler observations have been measured during research or regular monitoring cruises. The Ferrybox observations are made from a SMHI research vessel in the Baltic region and a commercial ship from Aalborg to Nuuk.

The majority of the in situ data are surface observations from the GTS network, which are concentrated in the southern part of the North Sea and Baltic Sea. The northern Baltic Sea is very sparse in observations as well as the northern part of the North Sea.

2.1.2. Temporal availability

The number of in situ observations is measured by using the number of spatial and temporal box averaging values. Two definitions of a box were chosen (Table 1): a ‘1 nm (nautical mile) by 1 nm by 1 h’ box for representing high frequency ocean processes and a ‘6 nm by 6 nm by 1 day’ box for representing synoptic processes. The number of satellite observations was subsequently counted for the box-averaged data.

Table 1 shows the number of the in situ SST observations. The majority of the in situ SST observations are from VOS, fixed platforms and Ferrybox. For
the hourly-averaged 1 nm × 1 nm grid, the in situ system has 202,444 gridded data in total which is 555/day. Among them VOS and buoy observations make about 42.6% and 33%, respectively. For the daily 6 nm × 6 nm grid, the number of data is very much different. The total number of data is 50,420, which is about one fourth of the number of data counted in hourly 1 nm × 1 nm grid box. VOS data makes about 65.8% of the total gridded data while the number of observations from the fixed platforms was significantly reduced due to their high sampling frequency.

Near Real Time data accessibility is a key issue in operational oceanography. For the in situ SST observational network, only GTS and some of the buoy and Ferrybox measurements are disseminated in near real-time. Most of the measurements from national monitoring cruises (e.g., from CTD casts, TSG and undulating profilers) are not delivered in near real-time but with a delay from days to weeks. Operational exchange of these data should be further improved so that these data can be widely shared and used in real-time in the operational forecasting.

2.1.3. System error

The monitoring system error and sampling error of the in situ SST observations are two complicated but key issues in the studies which combine satellite and in situ observations. The system error is defined as the difference between the raw observations and the ground truth and it includes instrument error, error in data transmission and pre-processing etc. This section discusses only the system error while the sampling error will be dealt in Section 2.3. Emery et al. (2001) carried out a system error study for the GHRSST (Global High Resolution SST) project. Three types of in situ observations were included in their study: moored buoys, drifting buoys and VOS. Due to lack of a ground truth SST for comparison, the error statistics of the moored and drifting buoys and VOS observations were estimated only through inter-comparisons in averaged grid boxes of 50 km by 50 km. A similar approach will be used here to identify the system error of the VOS observations used in this study. Moored buoy measurements will be used as a reference to compare with the VOS data.

Before discussing the in situ data inter-comparison, it is useful to describe the instrument calibration procedure used in the monitoring activities for the different type of in situ instruments.

All the CTD casts used in this study were made in environment monitoring cruises in EU countries. These cruises are repeated sections several times a year. The instrument error is negligibly small (in an order of millidegree) because all the CTD instruments had been calibrated before used in the cruises. There could be some errors made in the data transmission and processing but this error is limited and will be further diminished by using the quality control (Section 2.1.4).

The undulating profiler and TSG observations were spatially continuous SST measurements provided by BSH (Bundesamt fur Seeschifffart und Hydrographie) and IOPAS (Institute of Oceanology, Polish Academy of Sciences). The BSH cruises with the undulating profiler and the TSG were made once a year, together with the regular German environment monitoring cruises. The TSG and the undulating profiler observations were calibrated later by using CTD casts. The IOPAS sections with the undulating profiler are also regular research cruises (4 times a year). Hence the instrument error of the SST data from the undulating profiler and TSG is assumed in the same order as the CTD cast.

All the buoys in Fig. 1 are maintained once a month. The buoy measurements are hence calibrated once a month by comparisons with the CTD casts made during the maintenance cruises. The instrument error of the moored buoy SST measurements is thus regarded as similar to the CTD data.

The system error in the VOS SST is a complicated issue. The VSOP-NA (Voluntary Observing Ship Special Observing Programme - North Atlantic) project (Kent and Taylor, 1991) was designed to quantify systematic errors in the VOS data. VOS SST can be measured by different instruments, e.g., a bucket sensor, a hull sensor or by using the engine room intake (ERI). Taylor et al. (1998) have re-examined the VSOP-NA results for SST. The authors found that the bucket and hull sensor data were in reasonable agreement at night, while the ERI data was comparatively warm. The hull contact data were less scattered than those from other methods. Using the hull
contact data as a reference showed that the ERI data were on average biased high by between 0.2 and 0.4 °C. The bucket values were possibly about 0.1 °C cold at night but became a warm bias by up to 0.4 °C as solar radiation increased. Normally the information of instrument type is not included in the GTS report. This is why it is very difficult to quantify the error of the VOS SST. However, the warm bias of the overall VOS SST was also found by Emery et al. (2001) by comparing the VOS data with the moored and drifting buoy observations.

As shown in Table 1, the VOS data takes account of more than 60% if counted in a 6 nm × 6 nm × 1 day grid box. It is therefore very important to identify its instrument error. Due to the high quality and relative large amount of the moored buoy SST data, the moored buoy SST observations are used in this study as a reference to identify the VOS measurement error. A grid box size of 3 nm × 3 nm × 6 h is used in the comparison, i.e., the data falling into the same box will be regarded as a part of the same data pair. The VOS data have been quality controlled (see Section 2.1.4), and the upper-most layer buoy measurements are used in the comparison, which vary from 0 m to 4 m depth. For the daytime, 10,599 buoy-VOS data pairs are used. It is found that the VOS is about 0.04 °C warmer than the buoy data; the VOS data has a standard deviation of 0.23 °C against the buoy SST. For the nighttime, 6615 data pairs are used. It is found that the VOS is about 0.05 °C warmer than the buoy data and the standard deviation is 0.27 °C.

It should be noticed that the sampling error is introduced into the VOS standard deviation estimations. Though the estimation of the sampling error will be described in Section 2.3.1, part of the sampling error estimation results is used here. A spatial sampling error estimation based on the quality checked Baltic-North Sea satellite SST in 2001 shows that, the-averaged sampling error in the Baltic-North Sea for a 3 nm × 3 nm box is about 0.13 °C. The temporal sampling error in a 6-h temporal box is estimated from half hourly buoy SST observations to be around 0.09 °C for daytime and 0.076 °C for nighttime. With these spatial and temporal sampling errors, the standard deviation of the VOS SST against the buoy SST can be corrected to 0.17 °C for the daytime observations and 0.22 °C for the nighttime observations.

2.1.4. Quality control

Most of the in situ observations are already pre-quality controlled from the data providers. One exception is the ferrybox data. Here, a stricter quality control procedure is applied. The first step in the quality control was to remove obviously erroneous data below −2 °C and above 30 °C. Listed below is the set of criteria that were used to quality control the data.

- QC-1: An anomaly was calculated from the pseudo climatology that was used with the satellite data. The standard deviation of all the anomalies were calculated (1.6 °C) and observations deviating more than 2.3 times the standard deviation from the mean, were discarded (2.6% failed).
- QC-2: An anomaly was calculated from satellite observations that were-averaged in 50 km and 5-day bins. The standard deviation of the differences was 0.96 °C. Observations more than 2.3 times the standard deviation away from the mean were discarded (2.8% failed).
- QC-3: The variability of the in situ observations were calculated in 50 km and 3-day bins. After graphical inspection, a maximum standard deviation of 1.5 was chosen. Observations in bins with larger variability were discarded (0.3% failed).

Note that an observation can fail an edit criterion more than once, and the total number of observations discarded in the above analysis is thus 5.5%. The reason for using several editing criteria was that they might capture different errors on the data. This is supported by the results where less than 0.3% of the observations failed more than one of the three criteria.

2.2. Satellite system

2.2.1. System description

In the Baltic–North Sea region, satellite SST data can at present be retrieved from both infrared sensors (e.g., NOAA AVHRR) as well as microwave sensors (e.g. Advanced Microwave Scanning Radiometer-EOS (AMSR-E)). Other infrared satellite sensors include AATSR (Advanced Along Track Scanning Radiometer), MSG (MeteoSat Second Generation – MSG) and MODIS (MOderate Resolution Imaging Spectroradiometer, from the satellites AQUA and TERRA). MSG has started sending SST (one snapshot every 15 min) since the end of January 2003, covering areas up to 60°N. However the quality of the MSG SST observations in areas near 60°N, such as Baltic–North Sea region, is still an open question. AATSR is still on-its-way to be available as a real-time product for operational agencies. MODIS is providing a 4.6-km resolution SST products but not in a real-time mode. All the infrared SST sensors are restricted by clouds. In general, AMSR-E SST can still be retrieved in cloudy area though rain rate may significantly influence the quality of the
AMSR-E SST retrieval. The coarse resolution (0.25°), the land contamination and the low precision (∼ 1 °C) limit the usage of AMSR-E data for high quality operational SST analysis in coastal-shelf seas. The NOAA AVHRR SST is thus still the major source of real-time operational SST products for coastal seas. SST data used in this study is from two satellite products in 2001: EUMETSAT Ocean and Sea Ice SAF (based on NOAA 14 and 16) and BSH (based on NOAA 12). The SAF products were retrieved using a nonlinear algorithm whereas the BSH products were retrieved using a more traditional multi-channel algorithm. Further details can be found in Høyer and She (2004).

2.2.2. Data availability and quality control

The official SAF product is available from 24 June, 2001 but additional data have been obtained for the first half of the year 2001 (S. Andersen, DMI- Danish Meteorological Institute, personal communication, 2003). The official SAF product includes day and night observations from 1 to 2 satellites (up to four sets of observations per 24 h) but the SAF data from the first half of the year are only nighttime data from one satellite. BSH provided two satellite products based upon the NOAA satellite 12, one for the North Sea and one for the Baltic Sea, which differs in spatial resolution. Daytime and nighttime observations are available for both BSH products during all 2001. The main characteristics of the satellite products are listed in Table 2. The regions covered by the satellite products are shown in Fig. 2.

The satellite data availability is also measured in the number of observations for the two categories of the averaged grid box, as for the in situ data. However, a quality control process has to be performed before the counting the data.

For SAF-2 data, a quality flag from 0 to 5 is associated with every pixel. The higher flag value, the better the quality of the observations (Brisson et al., 2001). The criteria used to determine the quality flags from 2 to 5 are mostly related to the probability of cloud contamination and the SAF data with different quality flags were therefore tested. It was decided to use data with quality flags 2 and 3.

### Table 2

<table>
<thead>
<tr>
<th>Product</th>
<th>NOAA satellites</th>
<th>Area</th>
<th>Period in 2001</th>
<th>Day/night (D/N)</th>
<th>Resolution (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAF-1</td>
<td>14</td>
<td>Baltic–North Sea</td>
<td>1 Jan–30 Jun.</td>
<td>N</td>
<td>2</td>
</tr>
<tr>
<td>BSH-b</td>
<td>12</td>
<td>Baltic Sea</td>
<td>1 Jan.–31 Dec.</td>
<td>D+N</td>
<td>1.2</td>
</tr>
<tr>
<td>BSH-n</td>
<td>12</td>
<td>North Sea</td>
<td>1 Jan.–31 Dec.</td>
<td>D+N</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Fig. 2. The regions covered by the different satellite products. SAF MNOR indicates the area of the MNOR SAF product, the FMHA-b is the BSH Baltic product and FMHA-n is the BSH North Sea product. The gray shading indicates the bathymetry.
flag 4 or 5 to minimize cloud effects. The SAF-1 data does not include quality flags for every pixel. The data that are missing probably correspond to a quality flag of 2 or lower in the SAF-2 data (S. Andersen, DMI, personal communication, 2003). To increase the quality of the data, a cloud erosion filter was applied, which removed the edge pixels around data gaps. The filter resulted in increased mean values as well as reduced noise, indicating that cloud-contaminated pixels were indeed discarded by the filter. No quality flags are available for the BSH data. A cloud-screening algorithm has been applied to the data in the processing. This probably corresponds to a pixel quality value of 3 in the SAF-2 data. To further increase the quality of these data, a cloud erosion filter as for SAF-1 was applied to the data. Finally, a climatology check was applied to the SAF and BSH products. More details of the above satellite SST quality control can be found in Høyer and She (2004).

By applying these above quality control processes, the SAF SST used in this study is actually only a subset of the original SAF products since low-quality observations were removed. It is expected this high-quality SAF subset has less data than the BSH products. Table 3 gives the daily data availability of both the BSH and SAF products (the subset). It is not a surprise to find that the number of SAF observations is about 20% less than the BSH products. The daily satellite observations for the 1 nm × 1 nm × 1 h grid box is more than a hundred thousands. In comparison with the 555 observations per day in the in situ observation system (Table 1), this means that the amount of in situ SST observations is negligibly small, and that the satellite SST dominates the coverage of the SST observational networks in the Baltic and North Sea. However, this does not necessarily mean that the in situ SST is not important. The quality of the in situ measurements is significantly higher than the satellite observations. The role of in situ and satellite SST observations has to be examined with more comprehensive methods.

### Table 3

<table>
<thead>
<tr>
<th>AVHRR</th>
<th>1 nm × 1 nm × 1 h box</th>
<th>6 nm × 6 nm × 1 day box</th>
<th>Number of effective days</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOAA</td>
<td>Day 71,542</td>
<td>3029</td>
<td>365</td>
</tr>
<tr>
<td>12</td>
<td>Night 82,152</td>
<td>3258</td>
<td>365</td>
</tr>
<tr>
<td>NOAA</td>
<td>Day 29,966</td>
<td>1276</td>
<td>134</td>
</tr>
<tr>
<td>14</td>
<td>Night 35,325</td>
<td>1381</td>
<td>309</td>
</tr>
<tr>
<td>NOAA</td>
<td>Day 30,291</td>
<td>1188</td>
<td>191</td>
</tr>
<tr>
<td>16</td>
<td>Night 30,043</td>
<td>1223</td>
<td>191</td>
</tr>
</tbody>
</table>

2.3. Quality of the satellite SST products

The ‘error’ of the satellite SST products here is defined as the error against the true value of a gridded ‘bulk SST’. The gridded ‘bulk SST’ is defined as a box-averaged SST. In practice, the gridded bulk SST is estimated as an averaged SST within a grid box of 6 nm × 6 nm in the horizontal, 5 m in the vertical direction and 4 h in time. The scale of such a box is typical for marine synoptic analysis and data assimilation. The gridded bulk SST can be derived from satellite and in situ SST observations. The quality of the satellite SST products is assessed by the difference between the gridded value of satellite SST and in situ SST. In addition to the instrument error of the in situ SST (as described in Section 2.1), the sampling error will be another major source of uncertainty in the satellite SST quality assessment. The absolute SST sampling error in a given grid box can be defined as the standard deviation to the mean SST in the box.

2.3.1. Sampling error of the gridded in situ bulk SST

The SST sampling error in a 4-dimension 6 nm × 6 nm × 5 m × 4 h box is governed by the variability (or noise level) in each dimension. The spatial sampling error in a horizontal 6 nm × 6 nm grid box is estimated from all the satellite SSTs used in this study. The annual-averaged sampling error is about 0.22 °C for the Baltic–North Sea. As CTD casts, fixed platforms and VOS are all local sampling instruments, the bulk SST estimated from these observations has an uncertainty of 0.22 °C in the horizontal sampling. On the other hand, the spatial continuous sampler such as the undulating profiler and TSG will have the least sampling error in horizontal directions.

For the sampling error in the vertical direction, Solvstean et al. (2003) found that, within the upper 5 m, there is no obvious relationship between the performance of the satellite SAF data and the depth of the in situ observations. For the analysis, 308,991 in situ observations have been collected for the period June 24th 2001 to August 2nd 2003. For every 1-m depth from 0 to 5 m, SAF products (pixel values) from NOAA 16 and NOAA 17 were directly compared with the in situ observations. In this case, the sampling error in the temporal and horizontal dimensions can be neglected. For daytime data, 3873 satellite-in situ data pairs were used. Assuming that ΔT is the value of the in situ SST minus satellite SST, for the upper 5 m, it was found that the vertical mean value of the daytime ΔT is −0.002 °C, with the maximum (0.08 °C) at 1-m depth and minimum (−0.06 °C) at 0-m and 4-m depths; the mean value of
the standard deviation of the $\Delta T$ is 0.75 °C, with the maximum value 0.85 °C at 2-m depth and minimum 0.60 °C at 3-m depth. For nighttime data 5926 satellite-in situ data pairs were used. For the upper 1–4-m water depth, it was found that the vertical mean value of the daytime $\Delta T$ is 0.185 °C, with the maximum (0.21 °C) at 3-m depth and minimum (0.17 °C) at 2-m depth; the mean value of the standard deviation of the $\Delta T$ is 0.47 °C, with the maximum value 0.57 °C at 2-m depth and minimum 0.36 °C at 1-m depth. Based on this study by Sølvsteen et al. (2003), the vertical sampling error can thus be estimated as less than 0.1 °C.

The sampling error in a 4-h temporal box can be estimated from buoy measurements with 30-minutes interval. The sampling error in a 4-h temporal box is estimated as 0.061 °C. In situ measurements from CTD casts and the GTS (from the VOS and part of the fixed platforms) contain the full temporal sampling error. However, high-frequency buoy and underway measurements can largely diminish this sampling error.

### 2.3.2. Pre-processing of satellite SST

In order to compare the satellite SST with in situ measurements, the satellite data have to be pre-processed. This includes matching SAF and BSH products and correcting satellite SST observations to the bulk SST.

It was found that there is an annual variation in the daily difference of nighttime BSH and SAF products. This annual variation of the daily SST difference may be caused by the different retrieval algorithms used in the SAF and BSH products. The BSH product is therefore corrected to match the SAF products by removing its bias with the annual variation. This makes SAF and BSH products comparable in the quality.

In order to be comparable with the in situ measurements, the satellite SSTs have to be corrected to the bulk SST which represents an averaged water temperature in the upper 5 m. Firstly a spatial average in a 6 nm × 6 nm grid box was made. Secondly a bias correction is applied to both the-averaged SAF and BSH SSTs, which was calculated individually for each satellite by comparison with a large amount of quality controlled in situ observations, described in Section 2.3.3. This is because SAF and BSH SSTs represent ‘skin’ temperature (i.e., temperature in upper 10 μm) while the in situ SST measurements are normally made at a few meters depth. A detailed description of the bulk correction can be found in Høyer and She (2007-this issue).

### 2.3.3. Comparison between satellite and in situ bulk SST

The corrected satellite nighttime bulk SST is compared with quality controlled in situ data. The RMS error of the satellite SAF nighttime data has the smallest RMS error of 0.57 °C. The RMS error of the BSH products is about 20% higher than the SAF SSTs. It was found that BSH products have a much lower quality than the SAF in the winter months when water temperature is lower than 5 °C. In the summertime, the difference of the RMS error between BSH and SAF products is less than 0.1 °C. The details can be found in Høyer and She (2007-this issue). It was also found that the daytime error is about 10–20% larger than the nighttime error.

It should be noted that these numbers also consist of instrument and sampling errors in the in situ SST, as discussed in Sections 2.2 and 2.3.1. Since more than half of the in situ observations are VOS observations, the RMS error in Table 4a can be over-estimated, due to significant instrument and sampling error in the VOS measurements. The true bulk SST $T_{true}$ can be expressed as

$$T_{true} = T_i + \epsilon_i + \epsilon_b$$  \hspace{1cm} (1)

where $T_i$ is the in situ SST measurements, $\epsilon_i$ and $\epsilon_b$ are the system error and sampling error, respectively. Set the error of the corrected satellite bulk SST as $\epsilon_{sat}$, we have

$$\epsilon_{sat} = T_{sat} - T_{true}$$  \hspace{1cm} (2)

where $T_{sat}$ is the satellite bulk SST. Replacing $T_{true}$ in Eq. (2) by using Eq. (1), and assuming that the sampling error is independent with the instrument error, we finally have

$$<\epsilon_{sat}>^2 = <T_{sat} - T_i>^2 - <\epsilon_i>^2 - <\epsilon_b>^2$$  \hspace{1cm} (3)

where $<>$ denotes mathematical mean. In practice, $<T_{sat} - T_i>^2$ is the same as the RMS error given in Table 4a since the bias of $T_{sat}$ has been corrected. By

### Table 4a

<table>
<thead>
<tr>
<th>AVHRR</th>
<th>RMS error satellite vs. in situ (°C)</th>
<th>Corrected RMS error (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOAA 12</td>
<td>Day 0.75</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>Night 0.69</td>
<td>0.60</td>
</tr>
<tr>
<td>NOAA 14</td>
<td>Day 0.67</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>Night 0.55</td>
<td>0.44</td>
</tr>
<tr>
<td>NOAA 16</td>
<td>Day 0.67</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>Night 0.58</td>
<td>0.48</td>
</tr>
</tbody>
</table>

The ‘Corrected RMS error’ means the RMS error is corrected using the system error and sampling error with Eq. (3).
assuming independency between horizontal, vertical and temporal sampling error, the sampling error in Eq. (3) can be written as

$$<e_{s2}^2> = <e_{s1}^2> + <e_{s2}^2> + <e_{s3}^2>$$  \hspace{1cm} (4)

where $e_{s1}$, $e_{s2}$ and $e_{s3}$ represent sampling error in horizontal, vertical and temporal dimension, respectively.

For the VOS observations, $(e_{s1}, e_{s2}, e_{s3}) \sim (0.22 ^\circ C, 0.1 ^\circ C, 0.061 ^\circ C)$; the system error $e_4$ is about 0.17 $^\circ C$ for the daytime data and 0.22 $^\circ C$ for nighttime; The RMS error of the satellite bulk SST can be corrected by using Eqs. (3) and (4). For the SAF nighttime data, the RMS error is reduced to 0.46 $^\circ C$ from 0.57 $^\circ C$ in Table 4a. This number is very close to what was obtained by Solvøsteen et al. (2003). This correction has been applied to all satellite day and night products. The results are also given in the Table 4a.

In order to look further into the RMS error of the satellite products, Table 4b gives the satellite error statistics related to different in situ platforms. Both SAF and BSH nighttime SSTs were used. In general all the platforms have a negative bias to the corrected satellite bulk SST except for the VOS, which is consistent with previous findings on the warmer bias in VOS observations (references). The underway measurements made by the towed profilers and TSG have the smallest standard deviation (0.28 $^\circ C$ for the towed profilers and 0.45 $^\circ C$ for the TSG) from the satellite bulk SST. This is because these data have negligible instrument and sampling errors. However, the results may not be representative since all the TSG measurements are from one cruise and the undulating profiler measurements are from 5 cruises. It is also noted that the Ferrybox has the highest standard deviation of 0.76 $^\circ C$ against the satellite SST. Since the Ferrybox data were not quality controlled, they will not be used in the following sections.

The standard deviations in the second column of Table 4b can be corrected using Eqs. (3) and (4). The results are shown in the third column of Table 4b. For the towed profiler and TSG measurements, both the system error and the sampling error are set to be zero. For the CTD and buoy measurements, only system error is set to be zero and the sampling error correction is applied. For the VOS data, a full correction is applied. It is interesting that the corrected standard deviations for the buoy, CTD and VOS data are very close, ranging from 0.53 $^\circ C$ to 0.58 $^\circ C$. We thus have obtained a consistent estimate of the quality of SAF and BSH nighttime bulk SSTs.

Processes such as the diurnal thermocline variability heat the upper meters of the ocean during the day (Robinson, 1985). This means that daytime satellite observations may not be as representative of the bulk SST as nighttime observations. Due to convective processes, nighttime observations are considered to give a better estimate of the bulk SST. In the following sections, if not specifically mentioned, the satellite SSTs used are quality controlled, box-averaged and ‘bulk’ corrected nighttime temperature observations. It should be noticed that the SAF nighttime data used in this study is just a high quality subset of the original SAF products because a large amount of data with a quality flag 3 are discarded. Hereafter the product is referred as ‘SAF subset’. The daytime satellite SST is discarded due to their lower quality. More details about the satellite SSTs can be found in Høyer and She (2004) and She and Høyer (2004).

### 3. Natural variability and characteristic scales

#### 3.1. Natural variability

For observational network evaluation, the natural variability (signal) and the uncertainty (noise, both in monitoring and natural variation) are two basic features to be examined. Fig. 3 displays the spatial distribution of the standard deviation of the annual/semi-annual signals and the SST anomaly in 2001. The annual and semi-annual harmonics are calculated from the box-averaged SAF subset and BSH nighttime products with a resolution of 6 nm×6 nm in space and 1 day in time. The SST anomalies are obtained by removing the annual and semi-annual harmonics from the box-averaged SST data. The standard deviation is estimated for both the annual and semi-annual variability as well as the SST anomaly in year 2001. As shown in Fig. 3 (left panel), the annual and semi-annual signals account for over 3/4 of the total variance. As the spatial and temporal scales of the annual and semi-annual harmonics are on the order of hundreds of kilometres and several months, the annual and semi-annual signals can be easily resolved.
by the existing SST observational network. Hence the major focus of this study is put on the SST anomaly. It is shown in Fig. 3 (right panel) that the standard deviation increases eastward from about 0.6 °C in the western North Sea to more than 2.0 °C in the eastern and northern Baltic Sea. Since the satellite SST observations have a measurement error of about 0.5 °C, this implies a very high noise–signal ratio in the western North Sea.

3.2. Characteristic scales

The characteristic scale in the ocean is a key index in assessing and designing observational networks. The basic idea is to understand the homogeneity of the characteristic scales and estimate the number of observations available (or should be) in a spatial–temporal box with the same measure of the characteristic scales. For a network assessment problem, such a study provides an effective coverage ratio which tells the percentage effectively covered by a given observational network. For an optimal network design problem, the spatial distribution of the characteristic scales is required for calculating the total number of observations required. The characteristic scale is often defined as a correlation scale in space and time (North and Nakamoto, 1989). Hence a spatial–temporal SST correlation model is required in order to calculate the characteristic scales. In this paper, an independent assumption is applied to the SST auto-correlations in longitudinal (\(x\)), latitudinal (\(y\)) and temporal (\(t\)) directions. The correlation model for the SST anomaly (with annual and semi-annual harmonics removed) can be described as

\[
 f(\Delta x, \Delta y, \Delta t) = e^{-(a^{*}\Delta x + b^{*}\Delta y + c^{*}\Delta t)}
\]

where \((\Delta x, \Delta y, \Delta t)\) represents the longitudinal, latitudinal and temporal lags, respectively; \(f\) is the spatial–temporal correlation function; the correlation coefficients \((a, b, c)\) and \((\alpha, \beta, \gamma)\) can be fitted from the satellite SST anomaly when assuming stationarity. In practice, the empirical correlations are calculated locally by using the SST anomaly averaged in 10 km × 10 km × 1 day box. The parameters were determined separately for the latitudinal, longitudinal and temporal correlations. The best temporal correlation model was obtained by calculating the auto-correlation of the time series in every grid point where more than 70 observations were available. The lagged correlations in space and time were calculated where more than 15 pairs were available. These minimum numbers were selected to ensure a robust result. The empirical correlations were averaged for the whole domain to obtain the coefficients \((\alpha, \beta, \gamma)\), which are (0.5, 0.5, 0.85), while coefficients \((a, b, c)\) are considered as locally dependent values. Similar functions have been used by, e.g. Gandin (1963) in meteorology and Roemmich (1983) in oceanography. Further details of the correlation model calculation can be found in Høyer and She (2007-this issue).
With the correlation model, a correlation length scale can be calculated for a prescribed de-correlation value. In previous studies (e.g., She and Nakamoto, 1996a,b), an e-folding scale has been used as the 3-dimensional characteristic scales \((L, M, N)\) are defined as

\[
f(L, M, N) = \frac{1}{e}.
\]

This corresponds to a single-dimension cutting correlation of around 0.7. Here we use a slightly lower cutting correlation, i.e. 0.6 in each single \(t, x\) and \(y\)-dimension. Fig. 4 shows the spatial–temporal scales when applying this definition. There exists a significant inhomogeneity. Spatial scales larger than 200 km are found in the southern North Sea (in both \(x\)- and \(y\)-directions) and in the eastern Baltic Sea (in \(y\)-direction only). Small scales (<100 km) in the northern North Sea may relate to high noise/signal levels. The largest temporal scales (>2.5 days) are found in the northern and eastern Baltic Sea whereas medium temporal scales (1.5–2.5 days) are seen in the southern

Fig. 4. Spatial distribution of characteristic scales in a) the temporal (unit in days), b) longitudinal (unit in kilometres) and c) latitudinal (unit in kilometres) directions.
North Sea and small scales (<1.5 days) in the western Baltic Sea, transition waters, northern North Sea and the Jutland coast.

4. Ad hoc SST observational networks

In order to answer the questions raised in Section 1, a variety of sampling schemes have been chosen for the assessment work (Table 5). The purposes for selecting these observational networks are also indicated in the first column of the table.

The daytime satellite observations are not comprehensively assessed in this paper. The only assessment experiment including the daytime satellite observations is the effective coverage assessment of both day and nighttime satellite observational network (referred as S3dn in Table 5).

<table>
<thead>
<tr>
<th>Objectives</th>
<th>Network abbrev.</th>
<th>Observing network</th>
<th>Experiment periods and types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assess the effective coverage of all satellite data</td>
<td>S3dn</td>
<td>All 3 satellites (both day and nighttime data)</td>
<td>Year 2001, effective coverage analysis applied.</td>
</tr>
<tr>
<td>Assess the impact of discarding in situ SST</td>
<td>S3</td>
<td>All 3 satellites (night time data)</td>
<td>Year 2001, effective coverage analysis, OI and OSE applied.</td>
</tr>
<tr>
<td>Assess the impacts of the spatial inhomogeneity of the existing SST observational network</td>
<td>SI</td>
<td>S3 + GTS + Buoy</td>
<td>Year 2001, effective coverage analysis, OI and OSE applied.</td>
</tr>
<tr>
<td>Assess the impacts of the data quality and assimilation method</td>
<td>S3c</td>
<td>The same as S3 except that NOAA 12 is not included during Dec., Jan.–Mar.</td>
<td>Year 2001, OSE applied.</td>
</tr>
<tr>
<td>Assess the impacts of SST from a varying number of satellites; and relative importance of data coverage and data quality</td>
<td>S12, S14, S2a, S2b, S3</td>
<td>NOAA 12, NOAA 14, NOAA, NOAA 12 + 14, NOAA 14 + 16, All 3 satellites</td>
<td>24 June–5, Nov. 2001, effective coverage analysis, OI and OSE applied.</td>
</tr>
</tbody>
</table>

The general quality of the existing satellite and in situ SST observational networks can be obtained by assessing the network SI — including the satellite and the major part of the in situ observations (i.e., GTS and buoy observations). The role of the in situ SST observations can be assessed by comparing the performance of the networks SI and S3 (i.e., with and without in situ observations).

A specific network S3c is designed for testing the impact of low-quality satellite observations and of the different data assimilation methods, though this is only a very preliminary study due to the scope of the paper (as shown in Sections 7 and 8). Høyer and She (2004) found that NOAA 12 has a poor quality when the temperature is lower than 5 °C. This is the reason why the NOAA 12 data in the four winter months were removed in the network S3c. A comparison between S3c and S3 will give an indication of the importance of bad satellite observations. For the OSEs in Section 7, the model initialisation used in S3c is different from the other observational networks, which also gives an indication of the impact of data assimilation method.

Finally, the gaps and the redundancy in using multiple NOAA satellites for deriving SST can be assessed by evaluating the networks with different number of satellites. As it was found that the contribution from the in situ data is negligibly small, no further in situ sub-networks (e.g., CTD casts, GTS, or buoy networks) are assessed in this paper.

A unified and independent in situ SST dataset is used for validating the SST products generated from the field reconstruction and OSEs. This dataset is a subset of the in situ observations where GTS and buoy measurements are excluded since they are used in the network SI. Further descriptions are found in Section 6.

5. Effective coverage analysis

5.1. Definition of effective coverage

Since the concept ‘effective coverage’ has not been widely used in the literature, it is necessary to give a clear definition of our understanding of the concept. The area effectively represented by a single measurement depends on the local spatial–temporal scales. The representative area of the measurement is proportional to the local characteristic scales. In order to quantify the representative area of a measurement, we have to define ‘the area effectively represented by the measurement’, i.e., effective coverage. In this paper, the spatial–temporal SST scales are used to define the effective coverage of a measurement, which is a spatial–temporal box centred at the location of the measurement and with
its sides equal to the spatial–temporal scales. Mathematically this is defined as follows: for a given grid cell \((x_0, y_0, t_0)\), if a grid cell \((x_i, y_i, t_i)\) satisfies

\[
f(x_i-x_0, y_i-y_0, t_i-t_0) \geq f_c
\]

(6)

Where \(f_c\) is the cutting correlation which defines the characteristic scales, the grid cell \((x_i, y_i, t_i)\) and \((x_0, y_0, t_0)\) is called a pair of ‘impact cells’. We use the \(e\)-folding scale in this study (i.e. \(f_c = 1/e\)), which means

\[
a(x_i, y_i)^\ast (x_i-x_0)^{0.5} + b(x_i, y_i)^\ast (y_i-y_0)^{0.5} + c(x_i, y_i)^\ast (t_i-t_0)^{0.85} \leq 1,
\]

(7)

The grid cell \((x_0, y_0, t_0)\) is regarded as being effectively covered either when an observation is made at this cell or when a number of impact cells are observed. In practice, if the grid cell \((x_0, y_0, t_0)\) is not observed, the grid cell is called ‘effectively covered’ only if there are four or more impact grid cells observed. For a given sampling scheme of an observational network, its effective coverage thus means the total area covered by the effectively covered grid cells. The ratio of the effectively covered grid cells to the total number of grid cells for a given period is called the ‘effective coverage rate’ of the sampling scheme. Similarly this ratio can also be calculated for each grid cell.

In practice, the effective coverage rate is calculated by using box-averaged satellite data with a resolution of 6 nm × 6 nm in space and 1 h in time.

5.2. Effective coverage of the different sampling schemes

The time series of the effective coverage rate for the different sampling schemes, which are defined in Table 5, is shown in Fig. 5. It is found that all sampling schemes have a low-frequency variation of the effective coverage rate in 2001, with a time scale of 2.5–3 months. The highest effective coverage is in May and other significant peaks are seen in August, October and December. The lowest effective data coverage is seen in November and January.

By using all day and night satellite SSTs from the 3 satellites (marked as S3dn — solid line in Fig. 5), an effective coverage rate of about 31% is obtained. If only nighttime data are used (S3 — dotted line in Fig. 5), the effective coverage rate is only about 20%. The observational networks from a single satellite give very different results: NOAA 14 (dashed line in Fig. 5) has 11% effective coverage rate but NOAA 12 shows much larger effective coverage rate. However, this is mainly caused by using different quality control procedure rather than using different sampling schemes, as shown in Section 2. The effective coverage rate of NOAA 12 (18%) is slightly different from that of all three NOAA satellites (S3), which means that the data redundancy of the satellite SST observations is high. The effective coverage rate of the sampling scheme SI (S3 + in situ) is almost the same as that of the S3; hence the time series plot of the SI is not shown here. This suggests that the contribution to the total effective coverage from the in situ SST observational network is negligibly small.

6. Field reconstruction error analysis

The daily SST maps are reconstructed using a multi-platform Optimal Interpolation (OI) scheme based on a space dependent spatial–temporal covariance model (Høyer and She, 2007–this issue). The method is applied to the SST anomaly datasets generated from the ad hoc SST sampling schemes described in Table 5.

6.1. Method description

The theory of OI has been known for many decades within meteorology (Gandin, 1963) as a robust method.

---

Fig. 5. Daily effective coverage rate (averaged in space) in 2001 for a variety of sampling schemes in Table 5. S12: NOAA satellite 12; S14: NOAA satellite 14; S3: all three NOAA satellites; S3dn: all three NOAA satellites with both daytime and nighttime sampling.
of filling gaps in sparse dataset such as sea level pressures or geopotential height. Within oceanography, more recent applications of OI have produced global 1° × 1° SST products with correlations in meridional and zonal directions (Reynolds and Smith, 1994). In this study the OI SST grid has a spatial resolution of 10 km and the scheme applies space dependent spatial (meridional, zonal) and temporal co-variances. In addition, the difference in noise level between satellite and in situ observations is incorporated in the method. The noise on the satellite observations were taken as the measurement error of (0.6 °C)². The mean in situ co-variances were used to derive an error variance for the in situ observations of (0.3 °C)². Both noise levels include both system error and sampling error. In practice, the inhomogeneous spatial–temporal covariance model is fitted from the SST anomaly dataset and the OI is performed locally. Further details of the OI method can be found in Høyer and She (2007-this issue).

6.2. Assessment of the SST field reconstruction error

The quality of an OI product can be evaluated by using the SST reconstruction error, which can be calculated in three different ways: 1) deriving the theoretical OI fitting error, which is supplied by the OI method, 2) validation against independent satellite data and 3) validation against independent in situ observations. For the satellite validation, 1/36 of the satellite observations were left out when reconstructing the SST field and subsequently used as independent data for validation. The in situ validation was performed against in situ observations from instrument types that were not included in the OI analysis e.g., CTD casts.

This section assesses the quality of a general OI products based on both satellite and in situ observation network. Since network SI (Table 5) covers all 3 satellites and major part of the in situ observations (i.e., GTS and buoy measurements), it is used to represent the overall SST observation network in the Baltic Sea and North Sea. Table 6 shows the SST field reconstruction error of the network SI based on the three validation methods described above.

6.3. Field reconstruction error for different observational networks

Fig. 6 shows the time series of the OI fitting error (against the independent satellite observations) for the ad hoc observational networks. For NOAA 14 (dashed line), climatology SST is used in November and December due to absence of the SST data from the satellite. As expected, the fitting error has an opposite seasonal behaviour compared to the effective coverage rate, i.e., a lower error in the summer months and a higher error in the winter months. Fig. 6 also shows a very small difference in the OI error (∼ 0.01 °C) when using NOAA 12 instead of all the 3 NOAA satellites (S3). This means that the data coverage quickly reaches its maximum with increased number of infrared sensors due to the cloud restriction.

Since the OI fitting error against satellite data is mainly controlled by the effective coverage rather than the quality of different types of satellite products, it is necessary to study the fitting error against the in situ

![Fig. 6. Daily OI fitting error (°C) in 2001 for a variety of SST observational networks. S12: NOAA satellite 12; S14: NOAA satellite 14; S3: all 3 NOAA satellites; S2b: SAF subset.](image-url)
Table 7
OI field reconstruction errors (validated against in situ observations) with different observational networks

<table>
<thead>
<tr>
<th>Networks</th>
<th>S14</th>
<th>S2b</th>
<th>S3</th>
<th>25% of S3</th>
<th>SI</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE (°C)</td>
<td>0.80</td>
<td>0.78</td>
<td>0.82</td>
<td>0.83</td>
<td>0.82</td>
</tr>
</tbody>
</table>

observations. Table 7 displays the annual-averaged OI fitting error of different types of observational networks with validation against the in situ observations. From this table one can examine the relative importance of data quality and effective coverage. In addition to using the networks described in Table 5, a new dataset is formed by extracting every fourth of the S3 observations (i.e., 25% of the S3 data), in order to examine the redundancy of the satellite data. The least OI fitting error (0.78 °C) was obtained by using the SAF subset (S2b). With both SAF and BSH products, it gives a higher fitting error (0.82 °C). By using 25% of the entire satellite dataset, only a marginal difference with the full set of the dataset was made. This suggests that, for the OI method, 1) the existing satellite SST has a large redundancy; and 2) the quality of the satellite SST is more important than the coverage. The negligible difference between the fitting error of S3 and SI means that the contribution from the in situ measurements is very small in the OI reconstruction.

7. Observing System Experiments (OSEs)

The quality of the SST observation products depends not only on the quality and coverage of the observations but also on the field reconstruction method used. The above OI method is based on a spatial–temporal covariance model, which does not contain model physics. The OSE makes use of both model physics and observations to reconstruct the SST field, and is thus a good supplement to the OI method, as the field reconstruction error should be largely reduced. In this section, the field reconstruction error is represented by the error of the ocean model nowcast products.

7.1. Model description

The OSEs are carried out by using a DMI operational ocean model BSHeom – 3D Circulation Model, developed by BSH – Bundesamt fuer Seeschifffahrt und Hydrographie (Dick et al., 2001). The model has three nesting levels with a focus on the Baltic–North Sea: a 2-dimensional (2D) Northeast Atlantic (NOAMOD) to provide surge boundary conditions for the Baltic–North Sea 3D circulation model (NBCMOD). Finally, a 3D coastal ocean model covering the German Bight and the Danish Straits (KUCMOD) is two-way nested in the Baltic-North Sea model (see Fig. 7). The model has been running operationally at BSH since mid-90s and at DMI since 2001. All three models are set up horizontally in spherical coordinates and vertically in z coordinate, with a horizontal resolution of 24, 6 and 1 nm. The model has 14 vertical layers. The top layer thickness is selected as 8 m in order to avoid tidal drying of the first layer in the English Channel.

The model is forced by hourly meteorological forcing (10 m winds, 2 m air temperature, mean sea level pressure, surface humidity and cloud cover) based
on DMI’s operational weather model HIRLAM (HIgh Resolution Limited Area Model). The forcing has a horizontal resolution of about 15 km. The surface heat flux is parameterised using bulk quantities of both atmosphere and sea or sea ice, respectively. The evaporation flux is taken into account only in the heat budget. Changes in water volume due to evaporation, precipitation and ice formation are ignored. The river run-off is given as daily-averaged data (a combination of measurements and hydrological simulations) in 2001. The vertical eddy viscosity is calculated by using a Richardson number dependent mixing length scheme. Further description of the model and model set up can be found in Larsen et al. (2007-this issue).

7.2. Assimilation scheme

The Kalman Filter (KF) provides an optimal linear unbiased estimate of the ocean state by using the model state and observations (e.g. Ghil and Malanotte-Rizzoli, 1991). The observations are assimilated sequentially as the deterministic model is integrated forward in time. However, due to computational limitations, a full KF is not usable for practical oceanographic applications. This is mainly due to a prohibitive cost of propagating the model error covariance matrix forward in time. It is therefore necessary to perform a number of approximations that reduces the computational cost of this propagation. The SST assimilation scheme used in this study is based on a simplified KF scheme, which was developed by Annan and Hargreaves (1999) for SST assimilation. A few assumptions were made in this scheme. The first one is that it is possible to adjust the temperature field without adjusting the other model fields. In the assimilation scheme we therefore only need to consider error covariance in the temperature field. Secondly, the error growth in the ocean mixed layer temperature field is mainly due to errors in the atmospheric forcing and in the bulk parameterisation, which is used to calculate the heat flux into the ocean. Thirdly, horizontal advection and vertical turbulent mixing in the ocean are expected to be the primary agents for redistribution of the error field. It is assumed that these processes act to distribute the error and do not cause the error to grow. Furthermore this method assumes zero horizontal covariance, perfect vertical correlation in the mixed layer and zero correlation between the mixed layer and layers below. More details on the data assimilation scheme can be found in Larsen et al. (2007-this issue).

7.3. OSE design

The ad hoc designed SST observational networks in Table 5 are used in the OSEs. The purposes of selecting these networks are also listed in Table 5. In total one control run and seven OSE runs are made. The control run is produced for year 2001 without SST assimilation. The seven OSE runs assimilate SST from the ad hoc networks listed in Table 5. Six of them use the simplified KF (Annan and Hargreaves, 1999). The remaining one assimilates SST from a specially arranged network S3c (Table 5), which is the same as the network S3 but without NOAA 12 data during December and January–March 2001. A slightly different OSE is applied to this network, where an OI pre-processing method has been applied to the SST observations before assimilation (details described in Larsen et al., 2007-this issue). In this way the horizontal correlation can be partly taken into account in the assimilation. As described in Section 4, the purpose of using S3c is to assess the impacts from low-quality SST observations in NOAA 12 in winter months and from a different initialisation scheme. Note that the OSE for the S3c is a very preliminary study on the impact of data quality and assimilation methods.

7.4. OSE results analysis

7.4.1. Annual-averaged model error statistics

The model error statistics are obtained by comparing the model nowcast SST and the independent in situ observations. The annually-averaged error statistics (based on about 40*10^3 independent model-observation data pairs) of the OSEs are given in Table 8 (for the four 1-year runs) and Table 9 (for the six runs during 25 June–5 November, 2001). It should be noted that two of the six runs (i.e., the control run and the S3 run) in Table 9 are directly taken from the corresponding 1-year runs. In Table 8, the control run has a positive bias of 0.78 °C (i.e., the model over-predicts the SST) and a RMS error of...
Fig. 8. Spatial distribution of model nowcast RMS error (units in °C). a): control run and b): OSE run by assimilating observations from the SAF and BSH products and in situ SST from GTS and buoys (i.e., observational network SI).
1.20 °C. With three satellites (S3), the bias is almost completely removed and the model RMSE is cut down by almost half. A very small improvement (0.01 °C in bias and 0.004 °C in RMSE) is seen when including the GTS and buoy measurements. By using the improved assimilation methods, an improvement of 0.02 °C both in bias and RMSE is achieved. It is noticed that the model RMSE is comparable with the measurement error of SAF+BSH SST (0.62 °C) and much better than the OI product. This suggests that the model dynamics has effectively reduced the noise part in the satellite measurements.

7.4.2. Spatial distribution of model error

Fig. 8 displays the spatial distribution of the model RMS error (averaged for year 2001) in the control run and in an OSE run for the network SI. It is seen that large errors (>1.5 °C) are found in the middle of the North Sea and in the middle of the Baltic Sea, in the control run. After assimilation, most of the areas with large RMS error disappear. Most of the North Sea, southern and mid-eastern Baltic Sea is covered by a small RMS error around 0.5 °C. The Baltic–North Sea transition waters and many parts of the Swedish coastal waters have relative large errors of 0.75–1.25 °C (Fig. 8b). This large error may be caused by several factors: 1) less satellite observations in coastal waters; 2) high-physical variability in transition waters and frequent upwelling events in the corresponding Swedish coastal waters and 3) relative high-model errors in describing the Baltic–North Sea water exchange (She et al., 2007-this issue) and upwelling events.

The northern Baltic Sea has a relative small error around 0.5–1.0 °C. However, since the amount of model-observation data pairs used in estimating the model error, the results in this region should be interpreted with caution.

7.4.3. Temporal variability of model error statistics

The moving average of the weekly RMS error from the OSEs and the control runs are shown in Fig. 9. Fig. 9a gives results of the four full-year runs. By comparing the control run in Fig. 9a with OI reconstruction error in Fig. 6, one can clearly see a lower error in winter months. This reflects that the contribution from the model physics is significant in wintertime. The model has the highest error in summer time.

![Fig. 9](image-url)
This temporal variation of the model error is also reflected in the OSE runs. All three OSE runs have the largest error reduction in April–November by a factor of about 50% due to high data coverage during this period. January has the smallest improvement both due to the poor data coverage, low quality of BSH SST (Høyer and January has the smallest improvement both due to the poor data coverage, low quality of BSH SST (Høyer and She, 2004) as well as relative small model error.

By comparing the RMS error of the OSEs with (SI, dashed line in Fig. 9a) and without in situ data (S3, dotted line in Fig. 9a), it is found that only a slight improvement (∼ 0.6% more of the RMS error reduction ratio) is made by using all the GTS and buoy measurements. This means satellite SST is dominant in generating the high-resolution SST gridded fields by using data assimilation, which is similar to the conclusion made in the statistical field reconstruction in the last section. This result is applicable for the entire period.

The results from the S3c OSE (the solid line in Fig. 9a) indicate the impacts of a slightly improved data assimilation method and the input data quality. When the same amount of satellite SST is assimilated during April–November, the improved assimilation method gives systematic better results. The results in January show the importance of removing low-quality data. A 0.15 °C RMS error reduction is gained in January when BSH data is removed. However the gain in February is not significant which may relate to the increasing model skill in this month. It becomes a significant loss in March when BSH data quality is becoming better and the data coverage takes an important role. The OSE of S3c also suggests that a better assimilation method reduces the requirement to the amount of observations.

The purpose of the OSEs in the Fig. 9b is to test the importance of data coverage by using a different number of satellites (as described in Table 5). In addition to the major conclusions obtained above, the temporal variation of the error reduction is revealed here. At the start of the OSE period, there is only a small difference among the OSEs using the BSH (S12) and SAF subset (S14+16). However in late September, the model RMS error difference between the S12 run and the SAF run becomes extremely large (more than 0.5 °C). The superior performance of the BSH products is due to its large data coverage. As can be derived from Table 3, the BSH products have about 20% more data per day than the SAF subset product.

8. Conclusions and discussions

A multi quality-indicator approach is used in this paper to assess the existing satellite and in situ SST observational networks in the Baltic Sea and North Sea. In this section we present major conclusions, remarks and recommendations of this study.

8.1. General assessment of the existing satellite SST observational networks

8.1.1. Concluding remarks

The quality of the SAF and BSH SST products is assessed by comparing with the quality controlled in situ observations. The system error and sampling error of the in situ SST are estimated and used to correct the estimation of the satellite SST accuracy. For the Baltic Sea and North Sea, the accuracy of the SAF high-quality subset (with quality flag 4 and 5) was estimated as 0.46 °C for the nighttime data and 0.58 °C for the daytime data. The accuracy of the BSH products was estimated as 0.60 °C for the nighttime data and 0.67 °C for the daytime data. The major quality difference between the BSH and SAF products were found in the winter months.

The amount of the SAF high quality subset SST is about 20% less than the BSH product. The effective coverage rate of the SAF+BSH, day and nighttime data is about 31% whereas it is 20% for nighttime data only. The effective coverage of the NOAA satellite 12 is around 18%, which is close to the effective coverage of all three NOAA satellites (12, 14 and 16).

Both OI and data assimilation methods are used in reconstructing gridded SST fields. The SST gridded product (hourly with 6 nm by 6 nm horizontal resolution) obtained with OI (by using nighttime SAF and BSH products) has a RMS error of 0.69 °C–0.82 °C by using different type of validation data. By using a 3D ocean model and weather forcing, a simple SST assimilation scheme and satellite observations around 20% effective data coverage (i.e., network S3 or S12), the overall quality of the Baltic-North Sea SST fields is high, with a bias less than 0.1 °C and RMS error of 0.64 °C.

However, as shown above, the error in the transition waters is still large. A further look at high-resolution (e.g., 2 km) gridded products in coastal region (especially upwelling areas) gave errors higher than the 6 nm × 6 nm resolution results. It is still an open question of the sufficiency of the existing SST observations in generating high-quality coastal water SST products.

The largest error of the assimilated SST product is shown in the summer months (∼ 1 °C) while smallest error is in Feb. Mar. Oct. and Nov. (∼ 0.5 °C). In its spatial distribution, a relative large RMS error of 0.75–1.25 °C is found in the Baltic-North Sea transition waters, and coastal waters with upwelling.
8.1.2. Discussions and recommendations

The above assessment of the quality of existing satellite observing system is rather conservative due to following reasons:

- The assessment of the quality of the SST gridded fields is method-dependent. It is noticed that the data assimilation method used in this study is not the best available, especially the horizontal correlation of SST data was not efficiently included. A better data assimilation scheme (e.g., Ensemble Kalman Filter, Evensen, 1994) may provide a better quality of the model SST products.
- The daytime satellite observing system is not comprehensively assessed. The daytime satellite SST has not been used in the OSEs and statistical field reconstruction experiments.
- There will be more high quality satellite data available in the near future (e.g. AATSR data from EU project Medspiration1). Use of AATSR data will certainly improve both the coverage and accuracy of the existing SST observational networks.

A recommended future research area is to improve the quality of the SST products in coastal region, transition waters and during summer months. This could include: improving the model physics for the focused space and time zone, using daytime satellite observations, using SST data from more satellites, optimally blend in situ and satellite observations in coastal waters and use in situ data to correct satellite observations.

8.2. Role of in situ measurements

8.2.1. Concluding remarks

The in situ SST used in this study consists of underway measurements from the towed profilers and TSG, CTD casts, ferrybox, fixed platforms and VOS. The VOS and fixed platforms take account of 75% of the entire SST measurements. The system error and sampling error are estimated. For the VOS measurements, the system error is 0.17 °C for the daytime data and 0.22 °C for the nighttime data. The system error for the CTD, fixed platforms, TSG and the towed profilers are much smaller than the VOS. The horizontal sampling error is 0.22 °C for a 6 nm × 6 nm grid box; the vertical sampling error in the upper 5 m is less than 0.1 °C; for a 6-h window, the temporal sampling error is about 0.075 °C for the nighttime data and 0.09 °C for the daytime data.

For the Baltic–North Sea in situ SST in 2001, there are 555 box (1 nm × 1 nm × 1 h)-averaged observations per day. This number is 123/day for a 6 nm × 6 nm × 1 day box. This amount is negligibly small comparing with the satellite data.

The OI and OSE experiments show that, as long as the satellite SST is used, the in situ SST has a negligible impact either on increasing the effective coverage or reducing the statistical and dynamical field reconstruction error.

8.2.2. Discussions and recommendations

1. The data assimilation scheme used in this study is not in favour of sparse sampling observational networks because the horizontal correlation has not been efficiently included. This is a limit of the study.
2. An OSE assimilating in situ satellite only, together with a more comprehensive assimilation scheme, may give an alternative view on the role of the in situ observational networks.
3. Reliable in situ measurements are required for the model and satellite SST validation. They can also be used to reduce the large scale measurement error in the satellite SST, which is the subject of on-going research but not mentioned in this paper (see Høyer and She, 2006).
4. A major obstacle for further improving the quality of the SST gridded products is the accuracy of the satellite SST measurements, especially in daytime. The OSEs show that the model results are improved by removing low-quality satellite SST measurements. More fixed platforms should be deployed (especially in the Baltic Sea) to improve the accuracy of the satellite SST.

8.3. Satellite data coverage and quality

8.3.1. Concluding remarks

It is found that when the number of infrared sensors increases, the data redundancy increases rapidly, which is due to the cloud limitation to the infrared sensors. For example, the effective coverage rate of NOAA 12, NOAA 12 + 14 and NOAA 12 + 14 + 16 networks is very much similar. From the effective coverage point of view, there is a large data redundancy in the existing SST satellite observational networks. This may also relate to the short interval (less than 4 h) between sampling times of the three satellites. Currently the SAF SST includes NOAA satellites 16 and 17, which has a larger sampling time difference. This may reduce the redundancy of the satellite data.

About 31% of the Baltic-North Sea area is effectively covered by the NOAA satellites 12, 14 and 16. An

1 http://www.soc.soton.ac.uk/lso/medspiration/.
alternative method to increase the effective coverage is to use microwave sensors, which is free from the cloud limit. The accuracy and resolution are two major obstacles to apply the current microwave SST in coastal seas.

Together with the field reconstruction method, the accuracy and coverage of the satellite SST are major factors governing the quality of final SST products. To study the relative importance of data accuracy and data coverage, an assessment is done for the two satellite products: one with larger coverage but lower quality (BSH products) and the other with lower coverage but high quality which is a subset of the SAF products. However, the results of this assessment depend very much on the reconstruction methods. For the OI method, the accuracy of the satellite SST is more important than the coverage. Even using 25% of the satellite data can generate a similar quality as using the full dataset. In this sense, SAF high-quality subset is better than the BSH products. For the simplified KF method, both accuracy and coverage are important. The OSEs in summer months show that the model performance of using BSH products is better than using the SAF product, which is attributed to the larger effective coverage in the BSH products.

8.3.2. Discussions

The effective coverage rate defined in this paper is a simple but efficient index in assessing and designing observational networks. To estimate the effective coverage of a given observational network, one needs to know the spatial distribution of the characteristic scales of the observed parameters. In this paper, a spatial–temporal correlation model has been fitted from the satellite SST nighttime products, and the e-folding scale is used in estimating the effective coverage rate. The basic assumption is that the error and noise in the satellite data are uncorrelated in space and time. The uncertainties in estimating the effective coverage from the satellite SST products can be identified by using a detailed analysis of the type of satellite error and noise, as well as their impacts on the spatial–temporal correlation model estimation. This is an on-going research by the authors. Part of the study has been reported (Høyer and She, 2006).

Through assessing the relative importance of the satellite data accuracy and coverage, it was found that the satellite data accuracy has a significant impact on the final field reconstruction error. However the level of the impact depends on the reconstruction method. For the OSEs, both model error and assimilation scheme error affect the impact level of the satellite data accuracy in reconstructing the SST field. Hence all major steps that can improve the accuracy of the satellite data should be in favoured in the future. One such effort is to blend the AATSR and AVHHR measurements (in the EU project Medspiration). Another recommended effort based on this study is to make more high-quality in situ observations, with a sparse spatial sampling and dense temporal sampling, which have been proved to be important in improving the quality of the satellite SST (Høyer and She, 2006), though their effective coverage is negligibly small.

Acknowledgements

This work is supported by the European Commission under the Fifth Framework Programme. The project is ODON (Optimal Design of Observational Networks, Contract No. EVK3-2002-00082). Many thanks are given to Peter Loewe and Gisela Tschersich at the BSH for making their satellite SST data available to us. The work by the SAF project team and the help from Søren Andersen are also greatly acknowledged. Contributors of in situ data include: Bundesamt fuer Seeschifffart und Hydrographie (BSH), Swedish Meteorological and Hydrological Institut (SMHI) and Phillip Axe, National Environmental Research Institute (DMU), International Council for Exploration of the Sea (ICES), Proudman Oceanographic Laboratory (POL), Royal Danish Administration of Navigation and Hydrography (RDANH), Institute of Oceanology Polish Academy of Sciences (IOPAS) and the Baltic Sea Experiment (BALTEX).

References


