DANISH METEOROLOGICAL INSTITUTE

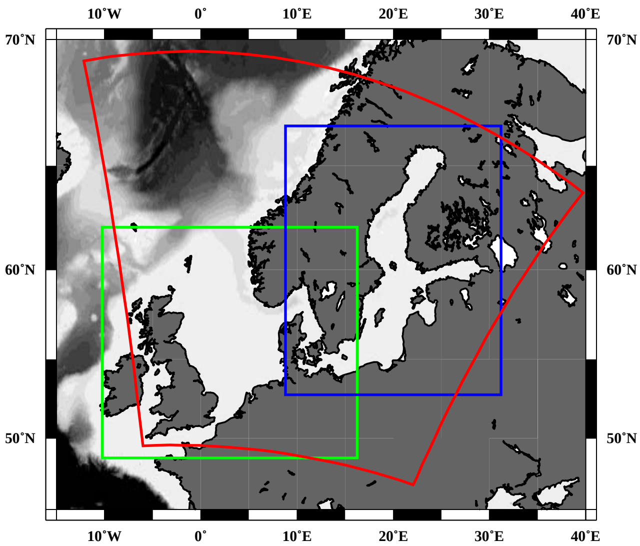
—— TECHNICAL REPORT ——

04-11

Validation of satellite SST products for the North Sea–Baltic Sea region

Jacob L. Høyer

Jun She



Copenhagen 2004

ISSN 0906-897X (printed)
ISSN 1399-1388 (online)

Validation of satellite SST products for the North Sea–Baltic Sea region

Jacob L. Høyer (e-mail: jlh@dmi.dk)
Jun She (e-mail: js@dmi.dk)

Danish Meteorological Institute, Copenhagen, Denmark

Contents

List of Tables

List of Figures

1 Introduction

The analysis of sea surface temperatures (SST) presented here is part of the work in the EU 5th Framework Programme project: Optimal Design of Observational Networks (ODON). The project focuses on the year 2001 where an unprecedented amount of in situ observations has been collected. Some of the tasks of the ODON project are to provide an overview of the availability of SST data, to validate the satellite observations and to construct a best possible high resolution SST for year 2001. This report presents the results that are obtained within these tasks. The in situ observations are thus used to validate high resolution satellite SST products from 3 different NOAA satellites and to perform a skin–bulk temperature correction in the North Sea–Baltic Sea region.

The report is organized as follows: Chapter 2 gives an introduction to the satellite products that are used in the rest of the report. This includes coverage, satellite intercomparison and quality control. Chapter 3 presents the available in situ SST observations, including SST and temperature profiles from scientific cruises, ships of opportunity, fixed bouys and ferrybox. The procedures used to perform a quality check of these observations are also described. Chapter 4 describes the comparisons between coinciding satellite and in situ observations and how the differences depend on e.g., the type of data, day of the year and the averaging scales. Finally, the report will conclude with a summary of the results and some comments about the future developments within this field.

2 Satellite Data

2.1 Satellite products

Two satellite products are used in the SST database for the year 2001, these are:

- SAF SST
- BSH SST

where the SAF product is the Ocean and Sea Ice SAF product[1] and the BSH is the satellite product from 'Bundesamt für Seeschifffart und Hydrographie'[2], which is used in their weekly SST analysis. 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, personal communication, 2003*). The official SAF product includes day and night observations from 1–2 satellites (up four observations per 24 hours) but the SAF data from the first half of the year are only nighttime data from 1 satellite. BSH provided two satellite products based upon the NOAA satellite 12, one for the North Sea and one for the Baltic Sea. Daytime and nighttime observations are available for both BSH products during all 2001 but they differ in the spatial resolution. The main characteristics of the satellite products are shown in Table 1. The regions covered by the satellite products are shown in Figure 1. The SST observations are based on infrared

[1] http://www.meteorologie.eu.org/safo
[2] http://www.bsh.de/

Product	NOAA Sat.	Area	Obs. in 2001	Day/night	Res. (km)
SAF-1	14	North Sea + Baltic Sea	1 Jan – 30 June	n	2
SAF-2	14 + 16	North Sea + Baltic Sea	24 June – 31 Dec	d+n	2
BSH-b	12	Baltic Sea	1 Jan –31 Dec	d+n	1.2
BSH-n	12	North Sea	1 Jan –31 Dec	d+n	1.5

Table 1: Characteristics of the satellite products used in the report. The SAF products are SST mosaics, constructed from several passes whereas the BSH products are from every pass.

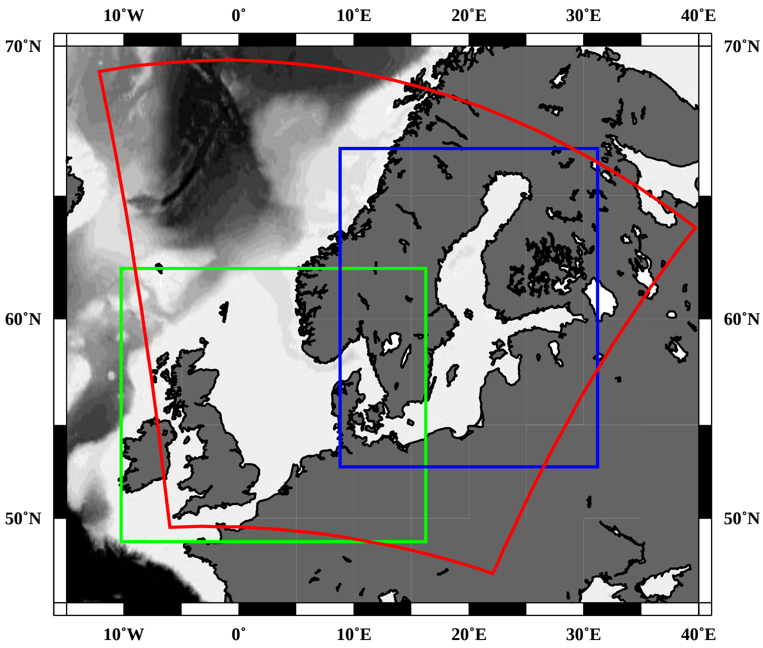


Figure 1: The regions covered by the different satellite product. Red line indicate the area of the MNOR SAF product, blue of the BSH Baltic product and green the area of the BSH North Sea product. Gray shading indicate bathymetry.

AVHRR observations on the polar orbiting NOAA satellites. The satellites are maintained in a sun-synchronous orbit, which means that they pass over the same geographical position at the same local time two times a day. Each satellite passes a location in the afternoon and in the night (early morning) and therefore provides a global coverage twice a day. Processes such as the diurnal thermocline may 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. The impact of this effect is examined in Figure 2, where day and night observations are compared. The differences are calculated when day and night observations are available within 24 hours and colocated within 20 km. The figure demonstrates that the differences vary throughout the year, with a maximum during summer of up to 1.5°C in the mean over the whole domain. This is in agreement with Kilpatrick et al. (2001). To eliminate this effect and to have a consistent satellite observation set, only nighttime satellite observations are used throughout the rest of this report, unless otherwise stated. This reduces the number of available observations with about 34% on a 5 km grid and daily averages.

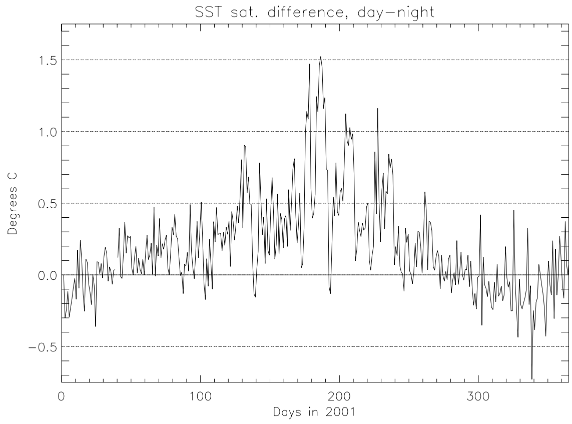


Figure 2: Mean difference between night and day satellite observations throughout the year 2001. Differences are calculated when observations are available in 20 km bins, both day and night within 24 hours.

Algorithms

The satellite products from SAF and BSH differ in the algorithms used to retrieve the SST from the measured brightness temperatures T. BSH uses the Multichannel algorithm based upon the observations from the AVHRR channels 4 and 5 to derive the SST

$$SST = A \cdot T_4 + (B_0 + B_1 \cdot (sec(\theta) - 1)) \cdot (T_4 - T_5) + C \cdot (sec(\theta) - 1) + D \qquad (1)$$

Where θ is the satellite zenith angle and A, B, C, and D are constants determined by validation and calibration against independent observations. The SAF project uses a more recent nonlinear algorithm that also includes a first guess SST value, T_{guess}

$$SST = A \cdot T_4 + (B_0 + B_1 \cdot (sec(\theta) - 1) + B_2 T_{guess}) \cdot (T_4 - T_5) + C \cdot (sec(\theta) - 1) + D \qquad (2)$$

where T_{guess} is taken as the climatologic SST value. The constants are determined separately for the four different products listed in Table 1. Spatial and temporal differences in the amount of in situ observations and in the reliability of the climatology may result in differences between the satellite products. The performance of the each satellite product and the error characteristics will be investigated in the following sections.

2.2 Satellite data coverage

For operational purposes, it is important to have an idea of the data coverage of satellite and in situ observations and how averaging in space and time will improve the coverage. The full coverage is calculated as all points where at least one observation in time is available throughout the year 2001. Gridpoints with sea ice are regarded as missing data. Figure 3 shows the spatial coverage of the satellite data, when averaged in 20 km grids and 3 days interval. The lower figures show that the North Sea (up to 61°N) has a higher average data return than the Baltic Sea, which experiences a large annual variation in the data return. This is probably due to the large seasonal signal in the ice extent in the Bay of Bothnia

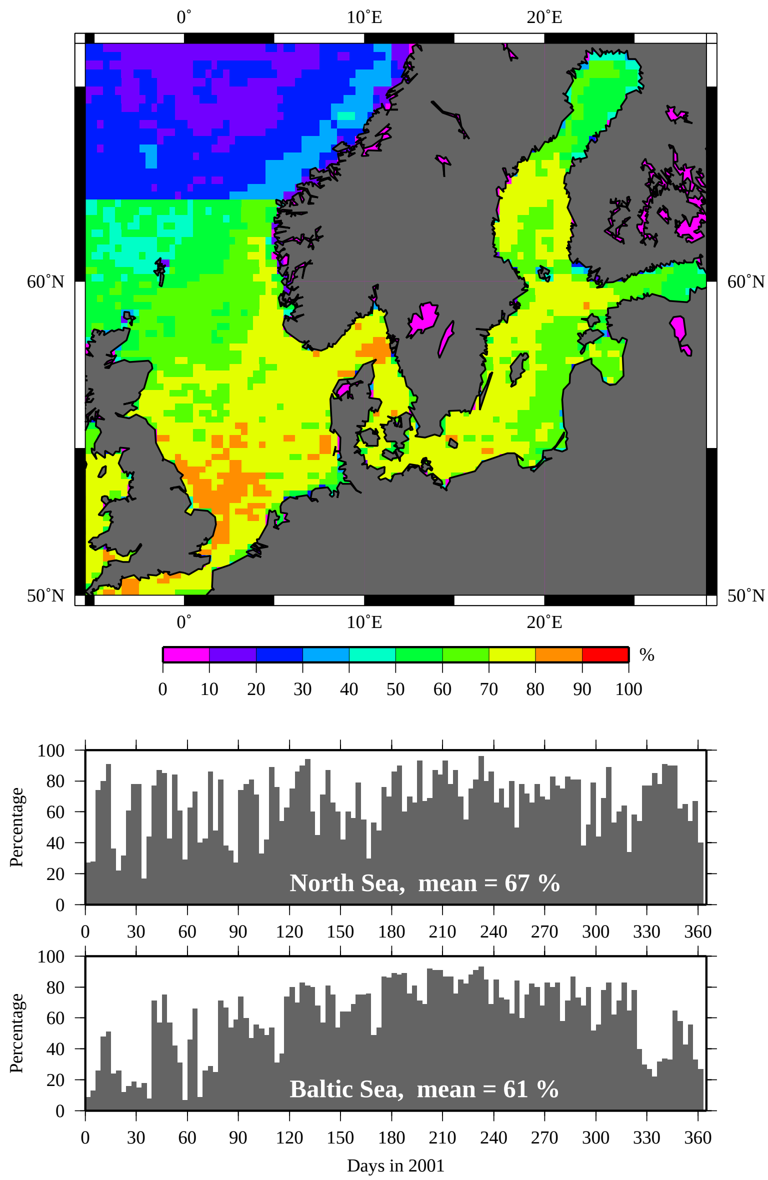


Figure 3: Top: Percentage of available satellite observations, when averaged in 20 km and 3 days. Below is shown the data coverage throughout year 2001 for the North Sea (50–61°N, 4°W–9 °E) and the Baltic Sea (54–66°N, 9–27°E).

and the Gulf of Finland. This is supported by Figure 4, where the coverage is shown during the 7 winter months (Nov-May) and the 5 icefree summer months (June-Oct). The impact of the ice cover is seen as a large difference between the two figures, where the coverage in the Bay of Bothnia and the Gulf of Finland changes from 20–30% during winter to around 80% during summer. The figure also shows that the highest data return is found in the Baltic Proper during the summer months.

Dependence on averaging scales

The coverage in Figure 3 and 4 depends on the spatial and temporal averaging scales. The relation is

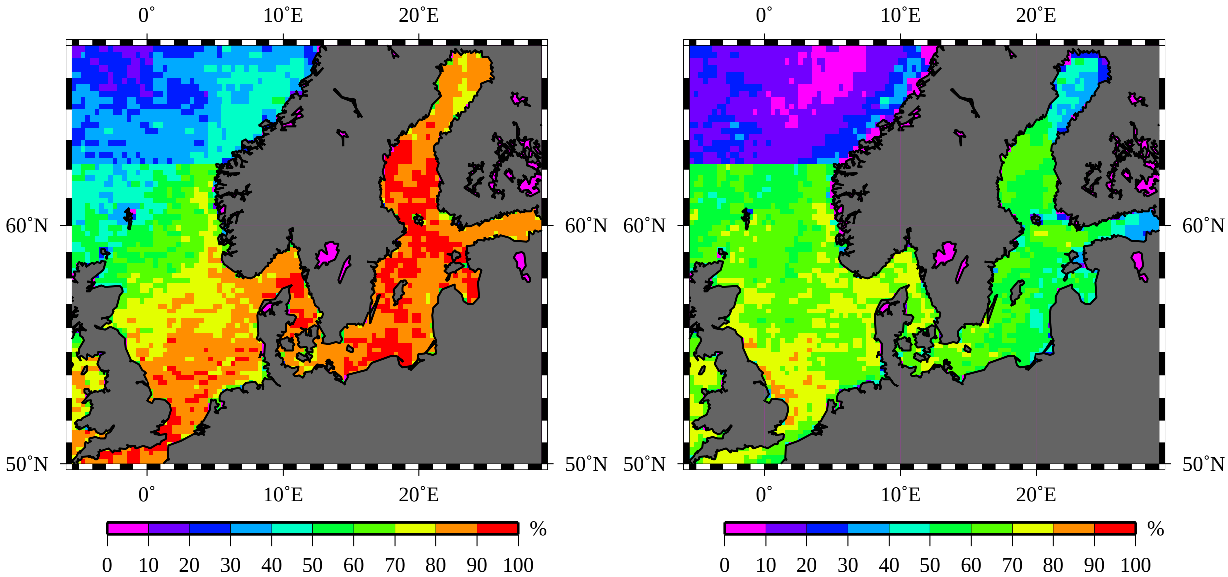


Figure 4: Left: The spatial coverage during summer (June 1–Nov 1). Right: The spatial coverage during winter (Nov 1–June 1). The spatial averaging is 20 km and the temporal averaging is 3 days.

examined in Figure 5 where the mean coverage is shown as a function of temporal (left) and spatial (right) averaging scales. Note that the absolute values can be too low due to the method used to calculate the coverage. The figure demonstrates that the largest gain in coverage is obtained with temporal averaging scales from 1–4 days and spatial averaging scales from 5-20 km.

2.3 Satellite data intercomparison

As described in section 2.1, the four satellite products have been processed individually, with varying quality control and algorithms. The consistency between the different satellite products can be revealed through a comparison of the different observations that are close in space and time. Averages over 20 km have been used to calculate the differences between observations taken during the same night. The histograms of the differences are shown in Figure 6 and Table 2 lists the associated statistics. The mean value and the width of the histograms indicate the relative bias and errors on the satellite observations. The observations may be separated by up to 20 km and 12 hours and there is thus a contribution from the natural variability to the errors. It is evident from Table 2 that the two SAF products overlap with 7 days only, which gives much less data pairs compared to the other products. The SAF-2 observations from the NOAA 16 satellite are in better agreement with the SAF-1 observations than the SAF-2 from the NOAA 14 satellite. This is somewhat strange since the SAF-1 are based on the NOAA 14 observations.

The comparisons between the BSH and the SAF-1 and SAF-2 reveal that while the SAF-1 is in general warmer than the BSH, the SAF-2 is significantly colder. This effect arise from a combination of a yearly variability in the BSH products and the two time periods for the SAF products (SAF-1 during first half of the year, SAF-2 during second half of the year). Figure 7 shows the daily differences between the

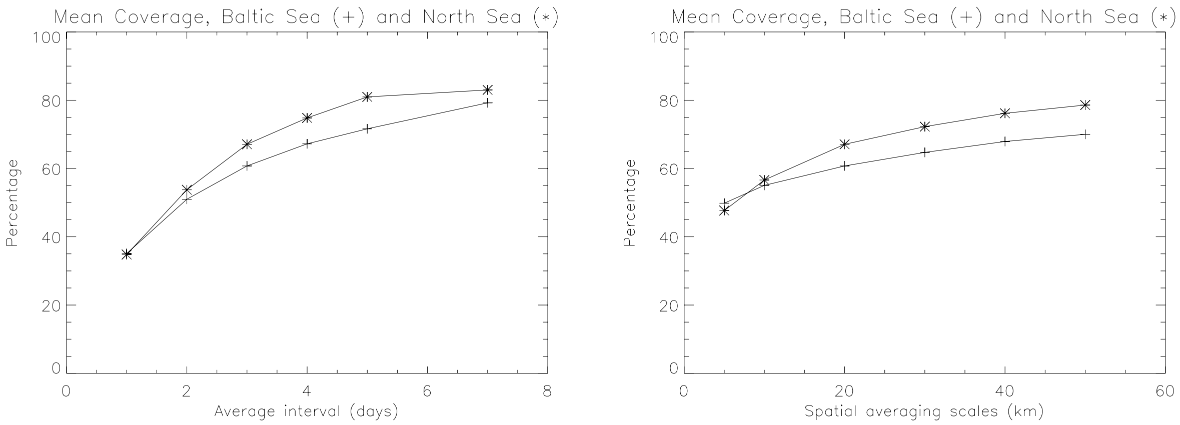


Figure 5: Mean coverage for the North Sea and the Baltic Sea as a function of temporal (left, spatial averaging = 20 km) and spatial (right, temporal averaging = 3 days).

Differences	Mean diff.	Number of obs.	Std dev.
BSH – SAF-1,2	-0.01	174592	0.61
BSH – SAF-1	-0.31	54730	0.57
BSH – SAF 2	0.13	123026	0.57
SAF-2(16) – SAF-2(14)	0.06	77176	0.53
SAF-1 – SAF-2(16)	0.002	3722	0.14
SAF-1 – SAF-2(14)	0.06	2819	0.34

Table 2: Statistics from the histograms in Figure 6.

BSH and SAF-1,2 products during the year for both the North Sea product and the Baltic Sea. A yearly variation is seen in both figures, with a variation of up to 2°C in the Baltic Sea. By comparison with in situ data, it was clear that the yearly variation was due to a variation in the errors on the BSH products. The yearly variation in the errors is correlated with the water vapor content in the atmosphere. It is likely to arise from errors in the B_0 and B_1 constants (equation 1) that perform the correction of the atmosphere. Furthermore, the yearly variation does not change much within different regions of the Baltic Sea or North Sea, which also point towards an error in the algorithms.

In order to generate consistency between the satellite products, the BSH SSTs were corrected every day, according to the smoothed red and green curves in Figure 7. Both the annual harmonic fit and the running mean were tested, with similar performances. The running mean resulted in slightly better results compared to in situ data and were therefore used to correct the BSH SSTs. In situ data could also have been used to correct the BSH data but the daily statistics were more robust when the SAF satellite data were used, with far more observations.

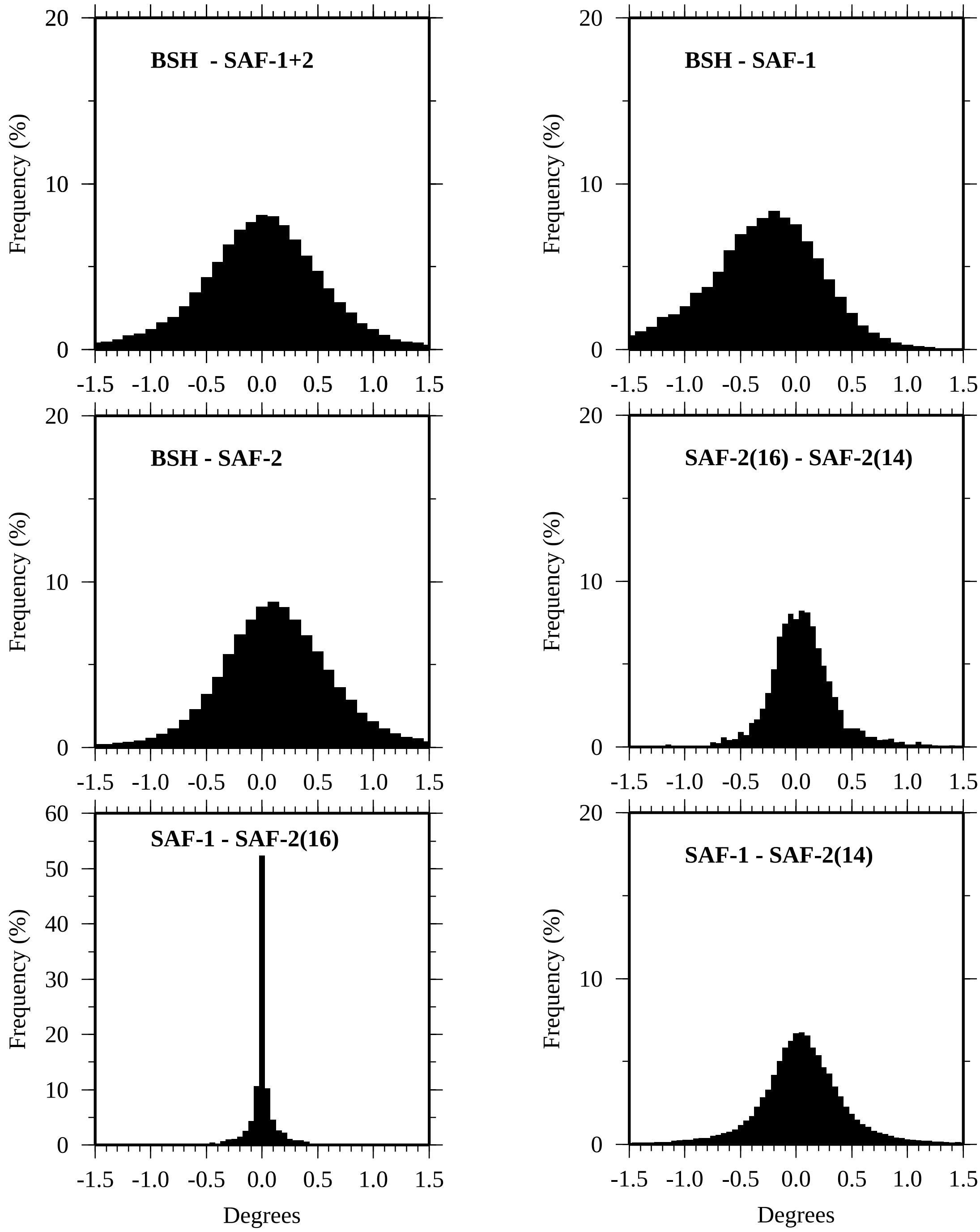


Figure 6: Histograms of the differences between the satellite products used here. All differences are calculated from coinciding 20 km averages where data were available during the same night. Numbers in parentheses mark if a specific NOAA satellite has been used.

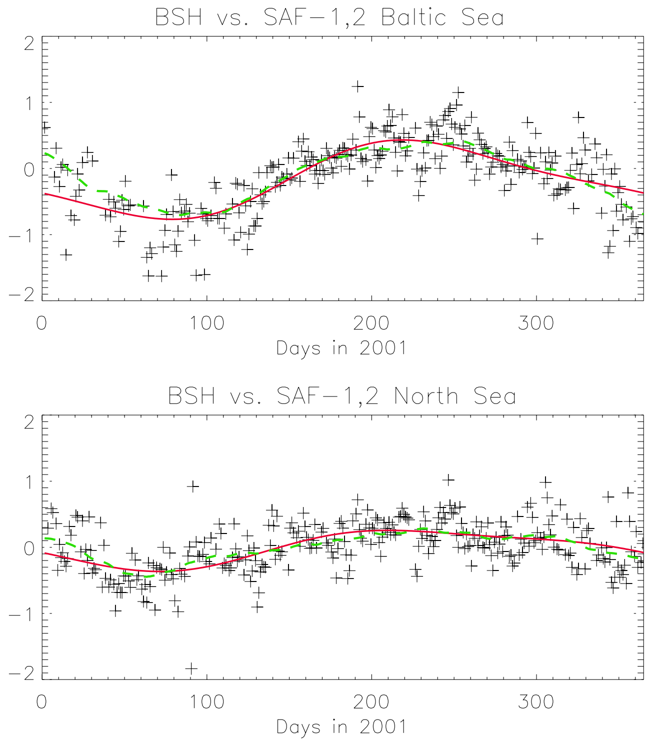


Figure 7: Daily differences between nighttime observations from BSH and SAF-1,2 for the Baltic Sea (top) and the North Sea (bottom). Differences are calculated where both data are available within 20 km. The red line mark the least square fit of an annual and semiannual harmonic and the green dashed line shows the running mean of the differences.

2.4 Quality Control

Several tests were performed to remove erroneous data and data contaminated by cloudy pixels.

SAF

A quality flag from 0 to 5 is associated with every pixel in the SAF-2 observations. The higher flag value, the better the quality of the observations and 0 means unprocessed where as 5 is excellent (Brisson and Marsouin, 2001). The tests used to determine the quality flag from 2–5 are mostly related to the probability of cloud contamination. SAF data with different quality flags were tested, and it was decided to use data with quality 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, 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, suggesting that cloud-contaminated pixels were indeed discarded by the filter.

BSH

No quality flags are available for the BSH data. A cloudscreening 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.

Climatology Check

A pseudo climatology was used to perform a quality check on all the satellite data used here. The climatology was constructed from daily SST fields from an operational ocean 3-D model run from Dec. 2001 to Dec. 2002. Jan-Dec 2001 were not available, and the 2002 observations were used instead. The mean difference between the model SST in 2002 and satellite observations from 2001 was subtracted the observations, to reduce the interannual differences. The temporal evolution of the mean SST was examined to be very similar in 2001 and 2002, suggesting that the pseudo climatology can be used to locate erroneous data.

As the pseudo climatology does not give an accurate estimate of the SST at a given time and position, the quality check can only be used to remove outliers which deviate substantially from the climatology. Based upon histograms of the residuals, satellite observations deviating more than -3.5°C or 4°C from the climatology were discarded. This removed less than 4% of the observations.

2.5 SST Variability

The variability of the nighttime satellite SSTs is shown in Figure 8 for 20 km and 1 days averaging scales. The significant east–west difference in variability, with increasing standard deviations towards

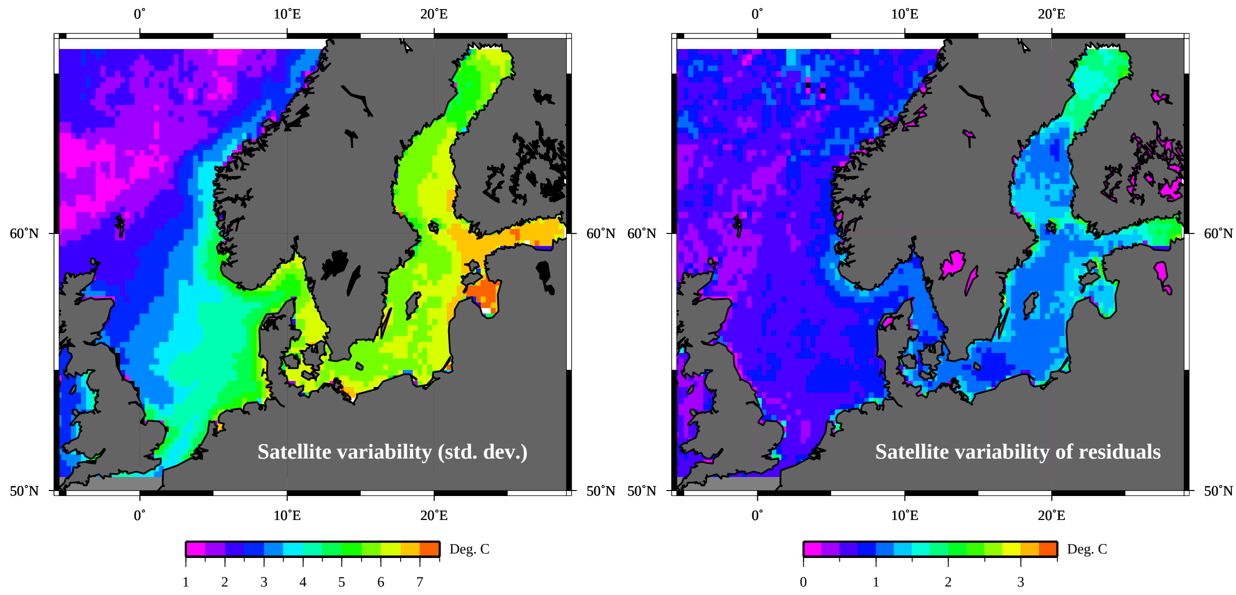


Figure 8: SST standard deviation of nighttime satellite SST observations, Left: Variability of the raw observations, right: variability of the observations where the annual and semiannual harmonics have been least square fitted and removed in every point. The data for both figures have been averaged in 20 km and 1 day.

the eastern Baltic reflects larger continental influence in the Baltic Sea. In addition, the permanent halocline in the Baltic Sea will also give higher variability compared to the North Sea where the waters are more homogeneous in the vertical. The right diagram in Figure 8 shows the variability of the satellite SST anomaly, where annual and semiannual harmonics have been fitted and removed at every point. The variability of the residuals has a spatial structure similar to the raw observations. The low variability level of the residuals indicates that a dominating part of the raw SST variability in this region is found in signals with annual and semi-annual periods.

3 In situ data

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. A check for redundant data was performed on all the available in situ observations to eliminate that the same observations were used twice. Two or more observations were defined to be redundant if they were within 0.01° in latitude and longitude, within half an hour in time and if, the temperature observations did not deviate more than 0.1°C. All the criterias above had to be satisfied before the redundant observations were discarded. About 5 percent of the observations were redundant, using this method. Table 3 lists the number of unique observations that has been obtained for 2001 from the different institutions.

Figure 9 shows the spatial distribution of all the available in situ observations of SST. The sea surface observations derive mainly from CTD profiles, bouy observations and from ships of opportunity. A ma-

Provider	Number of Obs.	Type of obs.	Num obs after QC	% bad data
DMU	1048	Profiles	966	8
SMHI	421	Profiles	384	9
ICES-CTD	1433	Profiles	1393	3
BSH-CTD	752	Profiles	652	13
IOPAS	4964	Profiles	4828	3
BALTEX	104	Profiles	53	49
BSH-TSG[a]	710	Surface	686	3
BSH-GTS	121057	Surface	113748	6
FERRYBOX	1545	Surface	1383	10
BSH-delphin	839	0-90 meters	839	0
MARNET[b]	33286	fixed depths	31674	5
RDANH[c]	18744	fixed depths	17527	6
Total	184903		174133	6

Table 3: List of the in situ observations of SST, provided by the different partners in the ODON project and others. Acronyms are: DMU: Danish Environmental Agency, SMHI: Swedish Meteorological and Hydrological Institute, BSH: Bundesamt für Seeschiffart und Hydrographie, RDANH: Royal Danish Administration of Navigation and Hydrography, IOPAS: Institute of Oceanology Polish Academy of Sciences, ICES: International Council for Exploration of the Sea, BALTEX: Baltic Sea Experiment. [a])36628 thermosalinograph observations were averaged into 710 hourly values, [b]) Observations from 6 bouys, [c]) Observations from 2 bouys. CTD is conductivity- temperature-depth measurements.

jority of the 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. No depth information is available for the GTS data but observations at different depths from bouy and CTD casts showed that the vertical variation in temperature of the upper $\sim$5 meters were in general small. This is in agreement with the results from Sølvsteen et al. (2003). Within the upper 5 meters, they found no obvious relationship between the performance of the satellite SAF data and the depth of the in situ observations. Apparently, depth information is not a crucial parameter for the in situ surface temperature observations and it will not be considered here.

As the in situ data originate from many different observation systems, the quality of the data varies substantially. Before using the data to correct the satellite observations, it is neccessary to perform a rigorous quality control which eliminates erroneous observations.

3.1 Quality Control

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 criterias 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 crit)

QC-2 An anomaly was calculated from satellite observations that were averaged in 50 km and 5-day

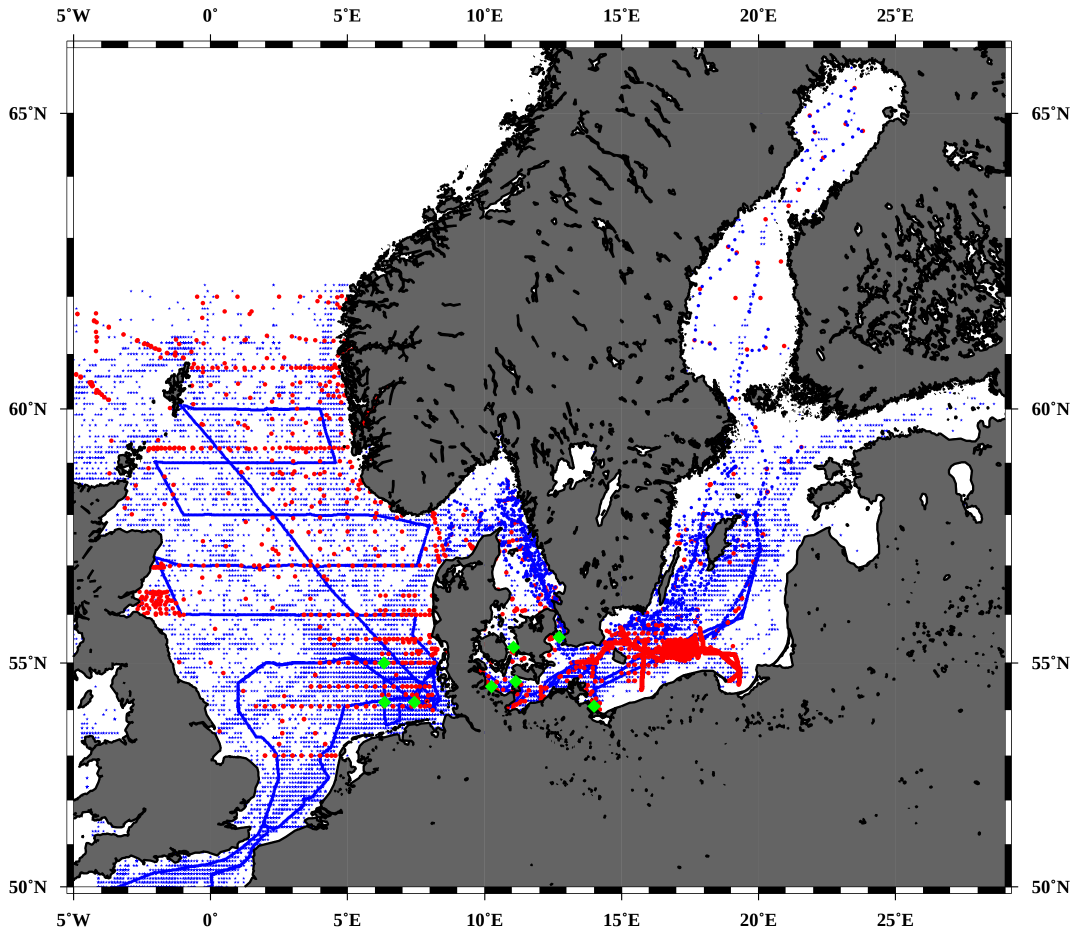


Figure 9: In situ observations of SST during 2001. Surface observations are marked with blue, profiles with red and bouys are in green.

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 crit)

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 were chosen. Observations in bins with larger variability were discarded. (0.3% failed crit)

Note that an observation can fail an edit criteria more than once, and total number of observations discarded in the above analysis is thus 5.5%. The limit of 2.3 in the standard deviation filter was selected to discard about 5 % of the data in total. The reason for using several editing criterias 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 criterias.

3.2 Coverage

The coverage of the in situ data is obviously much less than for the satellite observations. Figure 10 shows the spatial and temporal coverage of the in situ observations for the same averaging scales as for satellite data (Figure 3).

4 Satellite vs. In situ

This section presents the comparison between the quality controlled in situ observations described in section 3 and the adjusted and quality controlled satellite data described in section 2.

4.1 Coinciding observations ?

A pair of satellite and in situ SST observations are considered to be coinciding when they are within a certain distance in time and space (time and space window). In the comparison, the in situ observations are regarded as the true value, in space and time and a matching satellite observation is coinciding if it is located within a matchup time and distance. Note that the satellite observations are spatial averages of the available pixels within the spatial window whereas no temporal averaging is performed when comparing the observations. A specific time for each pixel is available for the SAF-2 observations, but the BSH product provides one time for each satellite pass, which is accurate to within 1 hour. The SAF-1 contains start time and end time information of the NOAA-14 mosaic, which is generated from several passes, typically three. As the satellite revolution time is 102 minutes, this means that the actual time for a pixel can deviate about 100 minutes from the middle time, which is used here.

The size of the spatial and temporal windows determines the number of comparisons that are available with the in situ data. Table 4 shows the number of available matchups (pairs of coinciding SST and in situ observations) as a function of window sizes. It is clear that the window sizes are crucial to the amount of data pairs that are coinciding.

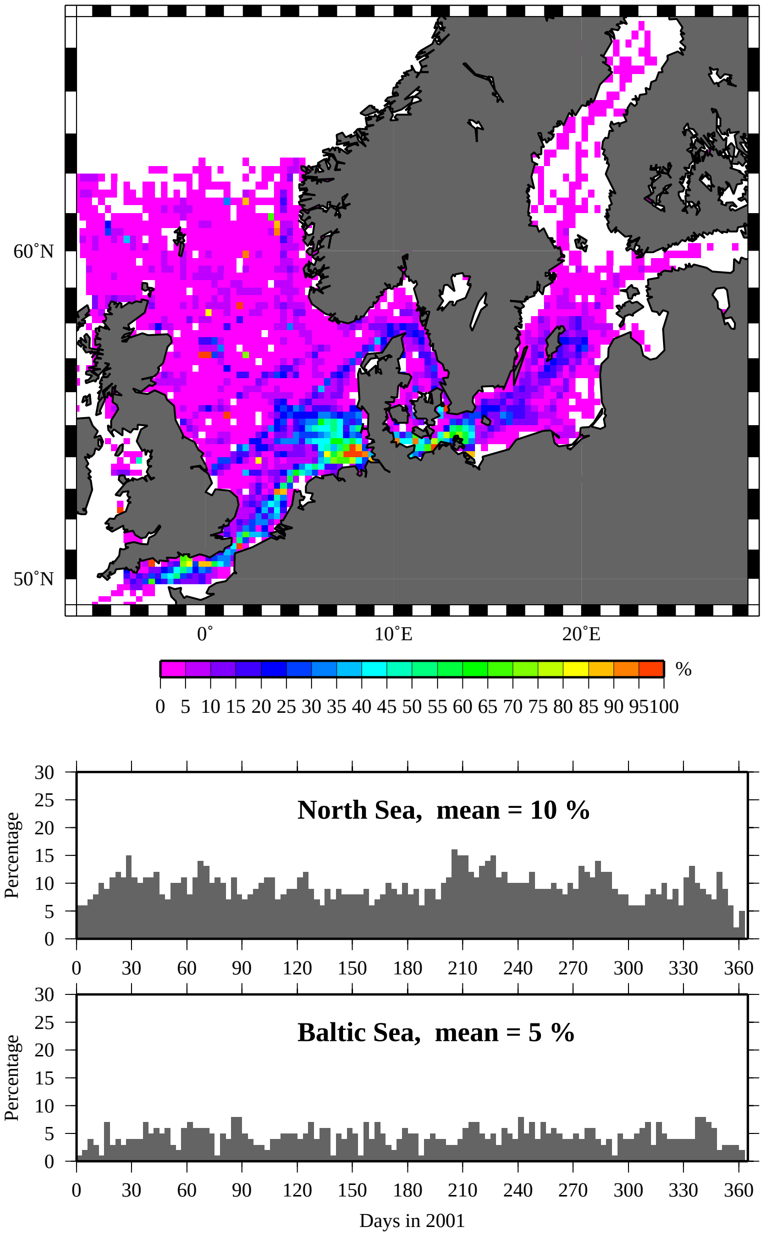


Figure 10: Top: Percentage of available in situ observations, when averaged in 20 km and 3 days. Below is shown the data coverage throughout year 2001 for the North Sea (50–61°N, 4°W–9 °E) and the Baltic Sea (54–66°N, 9–27°E).

Spatial window. km	Temporal window hours	Number of matchups
5	6	11032
10	6	14726
20	6	20550
10	4	9199
10	6	14726

Table 4: Number of available comparisons between in situ and satellite observations, as a function of the spatial and temporal matchup windows.

4.2 Error Statistics

A final quality check was performed specifically on the coinciding observations before the mean and standard deviations of the satellite and in situ comparisons were calculated. The quality check consisted of the application of a standard deviation filter to the residuals between climatology and satellite or in situ SST. A threshold of 2 standard deviations was chosen, discarding 4% of the satelite data, 5% of the in situ data and 7% of the pairs. This was applied for matchup windows of 10 km and 4 hours. Note that the matchup numbers given in Table 4 are after this final quality check.

The mean bias and standard deviation for comparisons between conciding in situ and satellite observations are shown in Table 5 for different spatial and temporal windows. The spatial averaging over several observations will lower the noise on the SST with a factor of $\sqrt{N}$ whereas the sampling error results in

Spatial window km	Temporal window hours	Mean diff °C	Std. dev °C
5	4	0.32	0.58
5	6	0.33	0.59
10	4	0.32	0.65
10	6	0.33	0.66
20	4	0.33	0.69
20	6	0.34	0.70

Table 5: Mean difference and standard deviation for satellite and in situ comparisons (in situ- sat) for different spatial and temporal averaging scales.

increased differences, when compared to in situ observations. Table 5 shows that the sampling error has the largest effect, as the standard deviations are reduced when the spatial window is decreased. The size of the temporal window of 4 or 6 hours does not have a large influence on the comparison with in situ data.

Effect of quality control
The quality control that has been performed on both types of data, is a crucial step when validating the satellite data. The effect of the quality control was investigated by calculating error statistics for different applications of quality control. The results are shown in Table 6 for matchup windows of 20 km and 4 hours. No quality control on any of the data result in an increase in standard deviation from 0.69 to

Quality control	Mean diff (°C)	Std. dev (°C)
Standard	0.33	0.69
None	0.43	0.84
Satellite data only	0.35	0.75
In situ data only	0.41	0.75

Table 6: Error statistics for different quality control of the satellite and in situ observations. The averaging scales are 20 km and 4 hours.

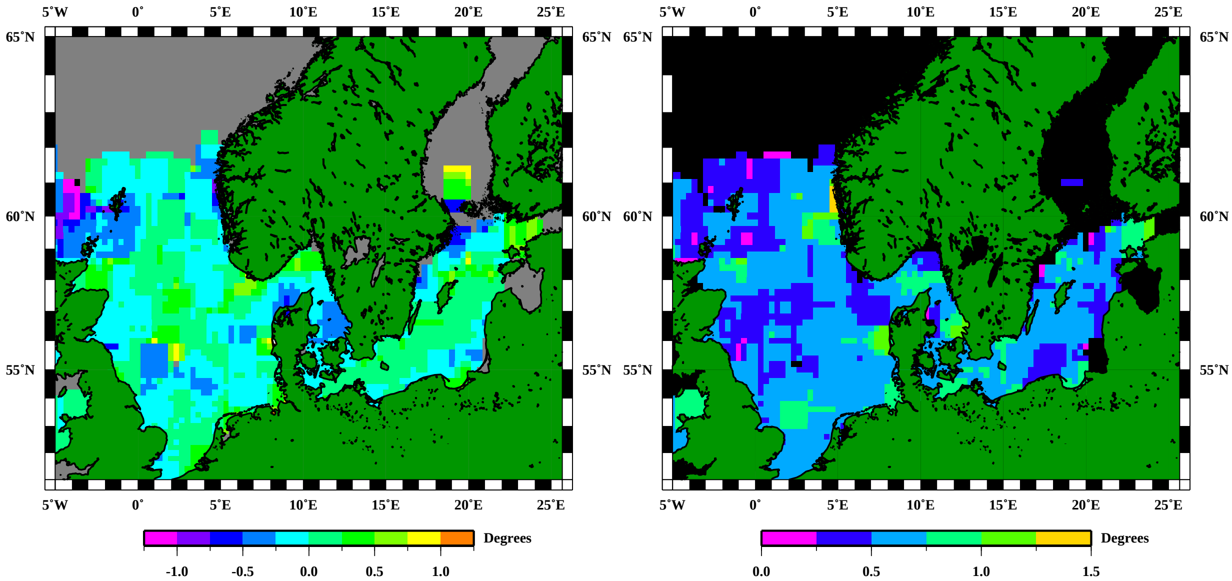


Figure 11: Spatial distribution of the mean difference (left) and standard deviations (right) between satellite and in situ observations. The matchup windows are 20 km and 6 hours and the values are calculated for spatial scales of 100×100 km.

0.84°C, whereas quality control on only one type of data gives an increase from 0.69 to 0.75°C. From the results where the quality control is applied to only one type of data, it can thus be concluded that it is equally important to check the quality of the satellite and the in situ observations. This indicates that the in situ data are not the truth, and that errors on the in situ observations are not neccessarily neglegible, as often assumed when compared to satellite data.

The spatial distribution of the mean bias and the standard deviations are shown in Figure 11 for the 20 km and 6 hour product. To obtain a more reliable result, each of the 20 km grid points represents averages over 100 km. The Danish waters appear to have a persistent cold bias, which means that the satellite data are too warm in this region. The standard deviations in the right figure show that by far the largest areas have standard deviations less than 0.75°C and only a few limited regions show larger variability.

Errors during the year 2001

Figure 12 shows the comparisons between satellite and in situ data as a function of the day in year 2001. The error statistics are relatively constant throughout the year with a clear tendency that the in situ observations are warmer than the satellite observations. A small annual variation (largest for the daytime

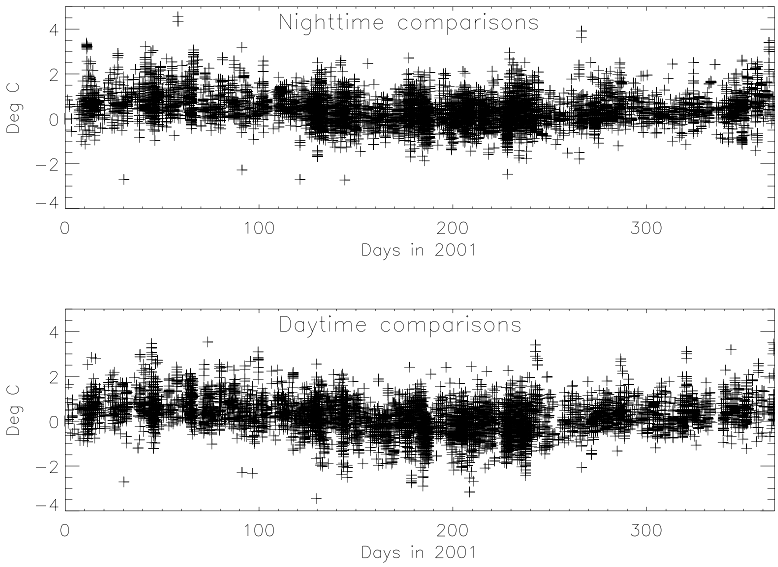


Figure 12: Comparisons between satellite and in situ SST observation for spatial and temporal windows of 10 km and 4 hours, respectively. Nighttime comparisons are shown in the upper diagram and daytime in the lower diagram.

observations) is found in the 24 hour mean difference, with larger offset during spring than summer. A further investigation of the differences revealed that the yearly variation is primarily due to the BSH products (largest for the Baltic) whereas the SAF products show a relatively constant bias throughout the year. This indicates that the BSH adjustments to the SAF satellite products did not remove all the annual variations in the BSH products. Improvements are thus required in determining the coefficients used in the satellite algorithmns in equation 1.

Errors of satellite products

As described in section 2.1, the satellite data consist of 4 different products. The mean difference and standard deviations for each of the satellite products are shown in Table 7 and 8 for nighttime and daytime observations, respectively. The scatterplots associated with the nighttime observations are shown in Figure 13.

Product	Mean diff (°C)	Std. dev (°C)	Number of Matchups
SAF-1	0.27	0.53	1047
SAF-2	0.12	0.58	2071
BSH-b	0.36	0.70	2695
BSH-n	0.42	0.67	3386

Table 7: Nighttime error statistics for the different satellite products, for averaging scales of 10 km and 4 hours.

The SAF data are seen to be of a better quality than the BSH data in terms of mean difference and standard deviations. The SAF-1 product from the first half of the year appears to be the most accurate product regarding standard deviation. However, the SAF-1 period from January to June is characterized

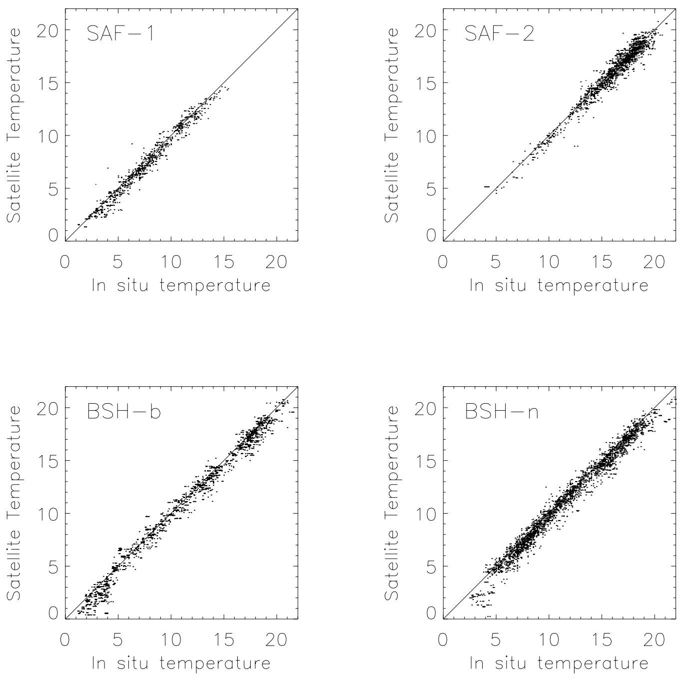


Figure 13: Scatterplot of satellite and in situ temperatures for the 4 different satellite products. The solid line indicates a perfect one to one relationship. The spatial and temporal windows are 10 km and 4 hours, respectively.

Product	Mean diff (°C)	Std. dev (°C)	Number of Matchups
SAF-1	NA	NA	NA
SAF-2	-0.09	0.67	1997
BSH-b	0.12	0.79	2204
BSH-n	0.09	0.69	2815

Table 8: Daytime error statistics for the different satellite products, for averaging scales of 10 km and 4 hours.

by a more homogeneous water column than from June to December (SAF–2 period) where a seasonal thermocline exists for most of the time. This may explain the higher accuracy of the SAF–1 data.

The better performance of the SAF data was anticipated due to the large effort the SAF project team put into validation and verification of these data. No routine validation is being done for the BSH Baltic product (*Gisela Tschersich, personal communication, 2003*), which explains the poor performance of this product. The higher noise on the daytime observations is in agreement with the generation of a diurnal thermocline as displayed in Figure 2, whereby the SST in the top layer becomes uncoupled from the SST at a few meters depth.

Errors for temperature bins

The scatterplot reveals that the mean bias and scatter around the perfect line is relatively constant for the SAF-1+2 data. The BSH satellite data show tendency towards underpredicting the in situ SST for cold and warm SST. This is supported by the results in figure 14, where the error statistics for 5°C temperature bins are shown. The SAF data display a relatively constant mean and standard deviation for all temperature bins where enough data are available to give reliable statistics. This is not the case for

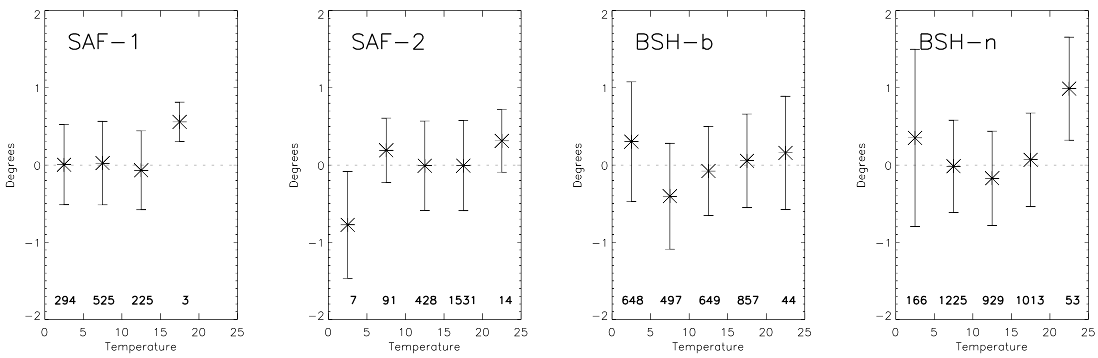


Figure 14: Error statistics, for temperature bins with a width of 5°C (in situ minus satellite). The asterisk indicate the mean difference for comparisons within each interval and error bars indicate the standard deviation. The bold numbers in the lower part of the figure, indicate the number of comparisons that are available within the temperature bin.

the BSH data where both mean difference and standard deviations vary for different temperature bins. The 0–5°C and 20–25°C SST bins are underpredicted by the BSH SSTs (in agreement with Figure 13) and these intervals also display the largest standard deviations. Note that all temperature intervals show higher standard deviations for the BSH data than for the SAF data.

Errors of in situ observations

The mean and standard deviation of the matchups within 10 km and 4 hour are shown in Table 9 for each type of in situ data. The quality of the in situ data varies significantly between the different types of observations. It appears that the SST from the towed profiling instruments (IOPAS and BSH-delphin) has a very low standard deviation. However, these data are from a limited number of cruises and in a small region of the area and no general statements can be made on that basis. Sampling issues are also a concern for the other types of observations where the number of matchups are too low to give reliable results. More data is therefore required to give conclusive statements about the performance of the different types of data. All data have been used in the following section to correct the satellite observations for the mean skin–bulk temperature difference

Provider	Type	Mean diff (°C)	Std. dev (°C)	Number of Matchups
DMU	Profiles	0.08	0.37	9
SMHI	Profiles	-0.33	0.38	14
ICES-CTD	Profiles	-0.39	0.49	47
BSH-CTD (TSP)	Profiles	0.26	1.00	29
IOPAS	Profiles	-0.05	0.35	343
BALTEX	Profiles	-0.34	0.02	2
BSH-TSG	Surface	-0.04	0.45	45
BSH-GTS	Surface	0.09	0.67	5634
Ferrybox	Surface	-0.13	0.76	44
BSH-delphin	0-90 meters	0.00	0.18	200
MARNET	Fixed depths	-0.19	0.51	1792
RDANH drogden	Fixed depths	-0.07	0.75	964
RDANH w26	Fixed depths	-0.41	0.72	76

Table 9: Error statistics for the different in situ observations, for averaging scales of 10 km and 4 hours.

Correction from skin to bulk SST

The temperature difference between the skin and bulk SST has been the topic of many investigations (Kearns et al., 2000, Emery et al., 2001, Murray et al., 2000). The differences have been observed to be higly correlated over large distances (Schluessel et al., 1990) and to depend upon heat flux and wind speed. Recent studies have attempted to parameterize the open ocean skin–bulk temperature differences by using wind speeds and net heat flux (Horrocks et al., 2003, Castro et al., 2003) but the studies have not yet been conducted for Shelf and Coastal seas. The lack of consistent parameterizations for areas like the North Sea–Baltic Sea determines that only in situ observations will be used here and no auxiliary data will be used to correct the satellite SSTs.

The skin–bulk correction is a tradeoff between the number of matchups that are available in space and time, and the complexity of the correction that can be performed. For the corrected SST product that is produced here, we have used the mean differences in Table 7 to perform a skin–bulk correction for the individual satellites. Each satellite is thus corrected by the mean temperature difference for the whole area and during all the year 2001. This conservative approach has been taken to avoid introducing unphysical temperature signals in the data, due to lack of representativeness of the matchups. Note that the adjustment of the BSH data to the SAF 1+2 data was performed on a daily basis and separately for the North Sea and the Baltic Sea. Some of the regional and temporal variations in the BSH skin–bulk differences may have been corrected in this adjustment. The accuracy of the final and corrected SST product is given in Table 10 for two different spatial averaging scales

Spatial window km	Temporal window hours	Std. dev °C	Number of matchups
5	4	0.58	6916
10	4	0.64	9199

Table 10: Accuracy of the final skin–bulk corrected satellite product for 5 and 10 km spatial averaging scales.

5 Conclusions and discussions

The following conclusions are reached in this report

- Satellite SST products from the SAFO&SI project and from BSH in Hamburg have been used together with a large amount of in situ data (~ 185000), to construct and validate a satellite SST product for year 2001.

- Only nighttime satellite data are used here, because the mean SST differences from day to night amount to 1.5°C, largest in summer.

- The satellite data coverage has large spatial variations, with largest data return in the southwestern North Sea and in Skagerrak. Averaging over 20 km and 3 days gives a mean coverage of 67% in the North Sea and 61% in the Baltic Sea.

- The in situ data coverage is largest in the German Bight and in the western Baltic. Averaging over 20 km and 3 days gives a mean coverage of 10% in the North Sea and 5% in the Baltic Sea.

- Quality control procedures have been developed and all satellite and in situ data have been quality checked. The decrease in the error due to the quality control is significant and equally important for in situ and satellite observations.

- The accuracy of the satellite product depends on the averaging scales. The larger matchup windows, the larger noise. For matchup windows of 10 km and 4 hours, the mean difference is 0.32°C with standard deviations of 0.65°C.

- The SAF data have higher accuracy (standard deviation about 0.55°C) than the BSH data (standard deviation about 0.7°C).

- Based on the comparison with in situ data, the satellite products are corrected individually for the skin–bulk difference. The accuracy of the final adjusted satellite product is 0.64°C standard deviation.

Other satellite products
The satellite data used for this study are from the NOAA satellites 12, 14 and 16. Other techniques and satellite systems exist that may be included in the future to provide a satellite SST product with better coverage. The AATSR instrument onboard the ENVISAT satellite is one of the additional data products that can be used. The instrument is very similar to the ATSR, which was flown at the European Remote Sensing Satellites, ERS–1 and ERS–2. The swath width of about 500 km gives a lower sampling compared to the AVHRR instruments but the measurement technique used by this passive infrared sensor is very advanced. A dual view technique uses two observations of the same point with different atmospheric path length to perform the atmospheric correction (Smith et al., 2001). This result in an ATSR (and expected AATSR) SST accuracy of about 0.3°C, which is superior to the AVHRR instruments onboard the NOAA satellites (Barton et al., 1995). Unfortunately, the near–real time data processing of the AATSR data is yet not as far as for the NOAA satellites, which is the main reason for excluding the data here.

Another possible satellite product, is the Advanced Microwave Scanning Radiometer (AMSR-E) instrument,which is flying on the Aqua satellite. In contrast to the satellites mentioned above, the AMSR–E instrument perform measurements in the microwave part of the spectrum and convert these to an observation of SST. The immediate advantage of the observations is that the measurements are not limited by cloud cover. Close to global coverage with a spatial resolution of 0.25° is obtained two times a day and the data are made available within hours of retrieval[3]. According to statistics on the web site, the precision of the data is about 0.6-1°C. These independence of cloud cover and the relatively good precision make these data attractive to use in the North Sea–Baltic Sea region. However, the spatial resolution of one of the satellite channels (6 GHz) is 74×43 km, which means that a very large landmask has to be used, to avoid contamination of near coastal observations. The effect of the landmask is seen in Figure 15, where observations are averaged from November 6–14 in 2003. The observations are not limited by cloud cover and the areas with missing data thus correspond to regions which are masked out by the landmask. The figure demonstrates the limitation of these data in near–costal regions like the North Sea and the Baltic Sea, where the areas with available data is greatly reduced, compared to infrared satellite observations.

6 Acknowledgements

The work in this report is funded by the EU project Optimal Design of Observational Networks (ODON), Contract No. EVK3-2001-00218) under the 5th Framework Programme. 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 für 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).

[3] see www.ssmi.com

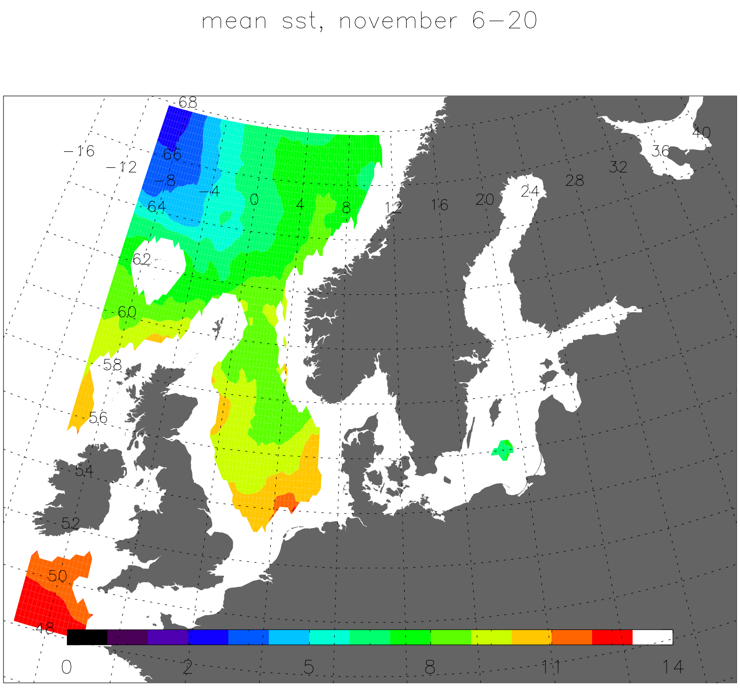


Figure 15: Mean SST from the AMSR-E microwave instrument on the Aqua satellite. The data are averaged over 14 days in November 2003 and the missing data corresponds to observations contaminated by land.

References

Barton, I. J., Prata, A. J., and Cechet, R. P. (1995). Validation of ATSR in Australian Waters. *J. Atm. Oce. Tech.*, 12(2):290–300.

Brisson, A., P. L. B. and Marsouin, A. (2001). *Ocean & Sea Ice SAF, North Atlantic Regional Sea Surface Temperature*. Meteo–France/DP/CMS, 22302 Lannion France. O&SI SAF Product Manual version 1.1.

Castro, S. L., Wick, G. A., and Emery, W. J. (2003). Further refinements to models for the bulk-skin sea surface temperature difference. *J. Geophys. Res.*, 108(C12). doi:10.1029/2002JC001641.

Emery, W. J., Castro, S., Wick, G. A., Schlussel, P., and Donlon, C. (2001). Estimating Sea Surface Temperature from Infrared Satellite and In Situ Temperature Data. *Bull. Meteorol. Soc.*, 82(12):2773–2785.

Horrocks, L. A., Candy, B., Nightingale, T. J., Saunders, R. W., O'Carroll, A., and Harris, A. R. (2003). Parameterizations of the ocean skin effect and implications for satellite–based measurement of sea–surface temperature. *J. Geophys. Res.*, 108(C3). doi:10.1029/2002JC001503.

Kearns, E. J., Hanafin, J. A., Evans, R. H., Minnett, P., and Brown, O. B. (2000). An Independent Assessment of Pathfinder AVHRR Sea Surface Temperature Accuracy Using the Marine Emitted Radiance Interferometer (MAERI). *Bull. Meteorol. Soc.*, 81(7):1525–1536.

Kilpatrick, K. A., Podest, G. P., and Evans, R. (2001). Overview of the NOAA/NASA advanced very high resolution radiometer Pathfinder algorithm for sea surface temperature and associated matchup database. *J. Geophys. Res.*, 106(C5):9179–9197.

Murray, M. J., Allen, M. R., Merchant, C. J., Harris, A. R., and Donlon, C. J. (2000). Direct Observations of Skin–Bulk SST variability. *Geophys. Res. Letter*, 27(8):1174–1174.

Robinson, I. S. (1985). *Satellite Oceanography, An Introduction for Oceanographers and Remote-sensing Scientists*. John Wiley & Sons.

Schluessel, P., Emery, W. J., Grassl, H., and Mammen, T. (1990). On the Bulk-Skin Temperature Difference and Its Impact on Satellite Remote Sensing of Sea Surface Temperature. *J. Geophys. Res.*, 95(C8):13341–13356.

Smith, D. L., Delderfield, J., Drummond, D., Edwards, T., Mutlow, C. T., Read, P. D., and Toplis, G. M. (2001). Calibration of the AATSR instrument. *Adv. Space Res.*, 28(1):31–39.

Sølvsteen, C., Krezel, A., and Löwe (2003). *Calibration of SST algorithm used for satellite data*. Royal Danish Administration of Navigation and Hydrography, Copenhagen. Report from WP3 in the BOOS SST group, version 2.