OPERATIONAL SST RETRIEVAL FROM METOP/AVHRR
Validation report

Version 2.0
July 2008

P. Le Borgne, G. Legendre, A. Marsouin and S. Péré
Météo-France/DP/CMS, 22307 Lannion, France
# Table of content

1. INTRODUCTION ........................................................................................................... 3
2. PROCESSING CHAIN ................................................................................................... 4
   2.1 Overview .................................................................................................................. 4
   2.2 Main processing steps ............................................................................................. 4
   2.3 Limitations of the version 0 chain under validation .................................................. 7
3. VALIDATION RESULTS AGAINST IN SITU MEASUREMENTS ................................. 8
   3.1 Global results .......................................................................................................... 8
   3.2 Regional results ...................................................................................................... 12
   3.3 Discussion and concluding remarks ........................................................................ 14
4. COMPARISON AGAINST AATSR SST ........................................................................... 15
5. VALIDATION BY USERS ............................................................................................... 19
   5.1 Controls by UKMO ................................................................................................. 19
   5.2 Controls by IFREMER ............................................................................................. 19
   5.3 Controls at DMI ...................................................................................................... 20
6. RESULTS OF THE CMS TEST CHAIN ........................................................................ 22
7. CONCLUSION ................................................................................................................. 24
   7.1 Summary of the results ............................................................................................ 24
   7.2 Modifications to the chain ...................................................................................... 24
   7.3 Related issues ......................................................................................................... 24
References ......................................................................................................................... 25
1. INTRODUCTION

The Centre de Météorologie Spatiale (CMS) of Météo-France has developed an operational processing chain to derive SST fields from the METOP Advanced Very High Resolution Radiometer (AVHRR) data. This work has been done in the frame of the Ocean and Sea Ice Satellite Application Facility (OSI SAF) of EUMETSAT. The OSI SAF is committed to produce:

- full resolution SST in satellite projection, corresponding to each granule of AVHRR data disseminated through EUMETCAST
- 12 hourly synthesis on a 0.05° global grid
- the 2km stereopolar NAR product, as a continuity to the present NOAA/AVHRR derived OSI SAF products.

It is the first time that SST is retrieved at the full AVHRR resolution over the globe. It is also the first time that the OSI SAF produces a global SST product. This required the development of a new processing chain in conditions of high data volume handling, processing and dissemination. This new chain also includes the use of fine scale global temperature and front climatologies. An attempt has been made to account for and, when possible, correct for dust aerosols.

This chain has been run on a routine basis since January 2007 with a progressive implementation of the elements. The Match up Data base, a key element for validation, was ready by March 2007. The development mode started with a stabilized version (version 0) in July 2007 for test by users and extensive validation by the development team. The METOP SST chain has been presented during the joint AMS EUMETSAT meeting in Amsterdam (Le Borgne et al, 2007A). A preliminary validation report has been produced in October 2007 (Le Borgne et al, 2007B), as well as a Product User Manual. The products have been declared “demonstrational” by EUMETSAT on the 14th of December 2007.

The version 0 METOP/AVHRR derived SST has been validated with the following means:

- building of a worldwide Matchup Data Base (MDB) gathering in situ measurements and coincident METOP/AVHRR data and analyze of the results,
- comparison of daily METOP SST fields with the corresponding ENVISAT/AATSR data (dual view retrievals, which are considered by the community as the “reference”),
- development and routine usage by operational experts of tools enabling rapid inspection of the SST products and the context of the calculations,
- implementation of a test chain to experiment alternative choices to the version 0 options,
- collection of feed back from users,
- study dedicated to high latitude problems, led by met.no.

This text is a more detailed report that replaces the preliminary validation report and is the CMS contribution to CDOP WP 21420 (METOP validation). It introduces briefly the processing chain that has been developed at CMS, with emphasis on the version 0 specificities. The next section presents and comments the validation results, obtained on a 11 month period, July 2007 to June 2008. It then synthesizes the feed back already available from key users (UK MetOffice, IFREMER, DMI). The results of a chain tested in parallel at CMS are then presented. This reports concludes by summarizing the results and describing the version 1 of the chain which is candidate for being operational. Note that this work has been presented during the IX GHRSST-PP meeting in Perros-Guirec, June 2008 (Marsouin et al, 2008).

A validation study led by met.no has been done over the high latitudes and is reported in a distinct document (Poulter and Eastwood, 2008); its results were also presented in the GHRSST-PP meeting ( Eastwood and Poulter, 2008).
2. PROCESSING CHAIN

2.1 Overview:

The CMS processing chain includes the following main steps (see figure 1):
- preprocessing: ingestion of the AVHRR L1C (L1B + cloud mask) data
- cloud mask control
- SST calculations
- proximity confidence value determination,

that will be detailed in the next section.

![Figure 1: Schematic diagram of the operational METOP SST chain](image)

2.2 Main processing steps

**Preprocessing:** This task delivers a “metagranule” that corresponds to a fixed time period (3 minutes). The SST processing chain ingests the original granules in AVH1C format which corresponds to the usual AVH1B to which the MAIA cloud mask has been added. MAIA is a threshold based cloud mask described in Lavanant, 2007 and includes a clear/cloudy flag and quality information at the pixel level. A workfile including the AVHL1C data and creating all the further requested variables is built for this metagranule.

**Cloud mask control:** A series of tests has been defined that consider various quantities such as the local values of gradient, temperature, probability of ice, etc.. For each test, a test indicator has been defined by comparison of the tested quantity (test_value) with a limit value (limit_value) and a critical value (critical_value). Outside this range of values either there is no problem, or the risk of errors is too high. The test indicator is defined as:

\[
\text{test\_indicator} = 100 \times \frac{\text{test\_value} - \text{limit\_value}}{\text{critical\_value} - \text{limit\_value}}
\]  

(1)

Indicator values below 0, or above 100 are assigned to 0 and 100, respectively. This approach enables the homogenisation of the test results on a unique scale: 0: no problem; 0-100: potential problem; 100: critical problem. The mask_indicator is initialized (0 = clear, 100 = cloudy) from the MAIA flag values. After initialisation, the value of mask_indicator is monitored for each clear water pixel (sea or lake). If a test results in a
critical problem identification (test_indicator =100), the mask_indicator takes also the critical value, the SST will take the missing value and the pixel will not be further considered. Hence the order of the tests matters. They are presented here below with the adopted hierarchy.

- **Primary cloud mask indicator:** derived from the MAIA quality information available.
- **Gradient indicator:** derived from the difference between the local 11 micron brightness temperatures gradients and the corresponding maximum climatological values calculated from the world Atlas of thermal fronts (see figure 2). The limit value of this quantity corresponds to the T11 noise equivalent gradient value. The critical value is a plausible margin which is reduced in the vicinity of cloud, so that for a pixel close to a cloud the critical value is more easily reached than far from clouds. The gradient indicator is calculated from the limit and critical values according to equation (1). In the present version, the gradient indicator is calculated by night only, because the use of visible channel allows in principle an efficient masking of the cloud edges and because diurnal warming may introduce local fronts that are not recorded in the climatology.
- **Stratospheric aerosol indicator:** in case of need, this indicator will record the intensity of Pinatubo like stratospheric aerosol phenomena. It will be calculated according to (1), where the test value is the aerosol optical depth.
- **Saharan dust indicator:** it is based on the use of the SEVIRI derived Saharan Dust Index (SDI, see Merchant et al., 2006) or the NAAPS Aerosols Optical Depth (AOD, see US NAVY, 2003) where the SEVIRI information is not available. Both information are in general consistent (figure 3).
- **Local temperature value indicator:** the calculated SST is compared to limit and critical values of the temperature deduced from the Casey world SST climatology (see above), by adding margins to the local value of the minimum SST climatology. These margins are function of the interannual standard deviation of the temperatures, of the distance to cloudiness and of the distance to coast (see figure 2).
- **Ice indicator:** derived from the probability of ice calculated by applying the met.no (Eastwood and Andersen, 2006) ice probability method, based on the use of the local value of the IR and visible AVHRR channels north (resp. south) of 50°N (resp. 50°S).

A synthesis of all the test indicators is made as the mean of all the meaningful test indicators. This mean value is the final value of the mask indicator. If an indicator is missing it is given a value =100 in the synthesis calculation. This synthesis is used, ultimately, to reflect the quality of the mask (see PCV determination below).

**SST calculation**

Two SST calculations are made: one in the context of cloud mask control, using the pixel values, and the final calculation using a smoothed atmospheric correction term over reliable data. There are many formalisms usable to derive SST from the AVHRR IR brightness temperatures. Based on previous experiences with GOES-East (Brisson et al., 2002) and SEVIRI (Le Borgne et al., 2006), the following algorithms have been selected:

\[
\text{NL : } \text{SST} = a T_{11} + (b T_{CLI} + c S_q) (T_{11} - T_{12}) + d + e S_q + \text{corr} \quad \text{daytime}
\]

\[
\text{T37_1 : SST} = (a + b S_q) T_{37} + (c + d S_q) (T_{11} - T_{12}) + e + f S_q + \text{corr} \quad \text{nighttime}
\]

\(T_{37}, T_{11}, T_{12}\) are the brightness temperatures at 3.7, 11 and 12 microns, respectively; \(\text{corr}\) is the correction term resulting from preliminary adjustment on the MDB; \(S_q = \sec(\theta) - 1, \theta\) is the satellite zenith angle and \(T_{CLI}\) is the mean climatological value.

Daytime is defined as in the solar zenith angle being in the range 0 to 90 degrees and nighttime in the range 110 to 180 degrees. In twilight conditions (solar zenith angle from 90 to 110 degrees), the SST is calculated through a weighted mean of daytime and nighttime algorithms.

The coefficients of the algorithms have been derived on a simulated brightness temperatures data base (see Francois et al., 2002). The prelaunch version of these coefficients were given in Le Borgne, 2006. The initial version of the coefficients are given in table 1. They are quite close to the prelaunch version but for the constant terms which have been adjusted on the first validation results (See Merchant and Le Borgne, 2004). A SDI correction term is calculated as a quadratic function of the SDI values (Merchant et al., 2006), for 0.1<SDI<0.8 (see figure3). This correction depends on the algorithm used. No corrections are made when there is no SEVIRI observations. In these conditions
there is a residual aerosol error that will be corrected, or flagged, in the future, using the NAAPS AOD (these residual errors may be apparent in the present data in the demonstration phase). At present the same algorithms are applied to retrieve surface temperature over sea and lakes.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
<th>corr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NL</td>
<td>0.99052</td>
<td>0.06641</td>
<td>1.16321</td>
<td>1.26512</td>
<td>0.16400</td>
<td>-</td>
<td>0.23</td>
</tr>
<tr>
<td>T37_1</td>
<td>1.01867</td>
<td>0.02109</td>
<td>0.68858</td>
<td>0.33056</td>
<td>1.02351</td>
<td>1.27303</td>
<td>0.13</td>
</tr>
</tbody>
</table>

Table 1: Coefficients of the non linear split window (NL) and triple window (T37_1) algorithms for METOP-A, with all temperatures expressed in Celsius.

Proximity confidence value (PCV) determination

PCV are dedicated on one hand to give the user a simple mean of filtering the data and on the other hand to partition the MDB in view of deriving the sensor specific error statistics (SSES) which are delivered for every SST calculation in the L2P/L3P format. It is essential here to adopt a method similar to or compatible with those of our partners in the GHRSST-PP project. For IR derived products, the normalised PCV scale shows 6 values: 0: unprocessed, 1: cloudy, 2: bad, 3: suspect, 4: acceptable, 5: excellent.

To calculate the PCVs, a representation of the risk factors on two axis has been adopted, one representing the cloud mask problem, the other the algorithm known risk of errors, on a scale from 0 (no risk) to 100 (critical risk), see figure 2.

Figure 2. Left: Schematic representation of the PCV attribution as a function of risks represented on a 2 axis system; right: example of PCV distribution.

In practice the first axis (X) is the mask indicator and the second (Y) is derived from the satellite zenith angle as Y=100 (θ/90). X<Xa and Y<Ya determines a PCV value of 5 (excellent) etc., as shown in figure 1. The values presently used in the processing scheme are: Xa= 20, Xb=30 and Xc=40; Ya, Yb, Yc correspond to satellite zenith angle values of 50, 60 and 70 degrees. Yc is off the limit of the AVHRR so practically no “bad” label is derived from the satellite zenith angle value.

Products fabrication

The processing is done at full resolution for each granule. At the end of the processing, a workfile is produced that contains the primary full resolution SST and all ingredients or intermediate products used throughout the processing: MAIA information, climatological values or ancillary data remapped at full resolution, intermediate indicators,... The delivered products are all derived from these workfiles, by selection of the adequate variables, and remapping on the corresponding grids.
2.3 Limitations of the version 0 chain under validation

The version 0 chain implemented in July 2007 has several known limitations that must be kept in mind when analyzing the validation results. These limitations are inherent to test versions that must keep dubious cases rather than being too restrictive. In the particular case of the version 0 chain, this led to the following:

1) In case of missing gradient climatological values, the gradient indicator will be assigned to 100, but the SST value is preserved

2) The limit and critical values of the Aerosol Optical Depth have been set to 0.5 and 1.5, respectively, which correspond to rather high loadings (which certainly affect the SST calculations)

3) The cloud/ice flagging developed by met.no has not been applied in twilight conditions (i.e. for sun zenith angle values between 85 and 95 degrees) where the results could have been uncertain.
3. VALIDATION RESULTS AGAINST IN SITU MEASUREMENTS

The operational validation of the METOP SST is based on the Matchup Data Base (MDB). The MDB collects in situ SST measurements from ship, moored or drifting buoys, available through the Global Telecommunication System (GTS) and the coincident full resolution satellite information, within 3 hours from the in situ measurement. The satellite information (calculated SST, brightness temperatures and reflectances) is extracted in a 21x21 pixel box centred on the measurement location providing the coverage of the box by clear pixels is larger than 10%. The MDB includes the in situ measurements (platform ID +SST + auxiliary measures) and all the variables of the workfile that results from the core of the processing.

The MDB is built with a 5 day delay to insure a good collection of the in situ data through the GTS.

The reference validation statistics are based on the exploitation of the MDB, as follows:
- Drifters only are considered
- Nighttime and daytime algorithms are considered separately
- To eliminate erroneous measurements, cases where the absolute value of the difference between the in situ measurement and the climatology exceeds 5 K are eliminated.
- The statistics are calculated from the differences between the central pixel of the validation box (when clear) and the buoy measurement.

The validation results are presented on a 11 month period, starting on 11 July 2007, with the routine delivery of the SST products in test the demonstrational mode. Throughout the period the algorithm and other chain characteristics have been kept unchanged. In this section, the term “error” always refers to the difference: satellite estimate minus buoy measurement.

3.1 Global results

Table 2 summarizes the statistics of the error and figure 2 shows a global map the mean error.

The distribution of errors as a function of the main parameters conditioning the SST calculations is presented in figure 2.

For nighttime cases, figure 4a shows limited variations of the bias with either the SST, the satellite zenith angle, the latitude or the longitude. The results are also stable in time. As expected, there is a clear dependency of the error characteristics with the mask indicator and consequently with the PCV. The clearer the validation box, the better the results, which illustrates the influence of the vicinity of clouds.

For daytime cases, the most striking feature in figure 4b is the negative bias for temperature above 30 Celsius or latitudes between 10 and 20 N. The two main factors explaining this deficit are failures in the Saharan dust correction and identification or limitations of the split window technique (see discussion below). For daytime and nighttime cases, the large positive biases observed over the Eastern Mediterranean Sea are due to erroneous buoy measurements. The buoys in question have been tracked during several weeks, identified, and reported to JCOMMOPS (the agency in charge of operational buoy quality control).

<table>
<thead>
<tr>
<th></th>
<th>All cases</th>
<th>2 “bad”</th>
<th>3 “suspect”</th>
<th>4 “acceptable”</th>
<th>5 “excellent”</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nighttime</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nbc</td>
<td>164478</td>
<td>4046</td>
<td>33693</td>
<td>59572</td>
<td>67167</td>
</tr>
<tr>
<td>bias</td>
<td>-0.01</td>
<td>-0.31</td>
<td>-0.11</td>
<td>-0.02</td>
<td>0.06</td>
</tr>
<tr>
<td>stdev</td>
<td>0.46</td>
<td>0.72</td>
<td>0.56</td>
<td>0.45</td>
<td>0.35</td>
</tr>
<tr>
<td><strong>Daytime</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nbc</td>
<td>229794</td>
<td>3353</td>
<td>32851</td>
<td>60697</td>
<td>132893</td>
</tr>
<tr>
<td>bias</td>
<td>0.12</td>
<td>-0.31</td>
<td>0.10</td>
<td>0.08</td>
<td>0.16</td>
</tr>
<tr>
<td>stdev</td>
<td>0.58</td>
<td>0.99</td>
<td>0.71</td>
<td>0.62</td>
<td>0.51</td>
</tr>
</tbody>
</table>

**Table 2**: Mean and standard deviation of the error, for all cases and as a function of the confidence level, from 11 July 2007 to 18 June 2008 (nbc is the number of cases)
Figure 3: Map of the mean error of the nighttime (top) and daytime (bottom) algorithm from 11 July 2007 to 18 June 2008; data are binned in 5 degree cells.
Figure 4a: nighttime error as a function of several parameters, from 11 July 2007 to 18 June 2008. Bias (+ and solid line) and standard deviation (squares and dashed line) are displayed as a function of: in situ SST, secant of the satellite zenith angle, latitude and longitude, PCV (confidence level), coverage of the validation box by clear pixels, mask indicator and time (from left to right and from top to bottom).
Figure 4b: Same as figure 4a but for daytime cases.
3.2 Regional results

The OSI SAF routine validation also produces nighttime and daytime error statistics calculated per oceanic basins which are those of the Met Office SST analysis OSTIA and which were proposed for regional studies within the GHRSSST-PP. The basins are the North and South Atlantic, North and South Pacific, Arctic, Southern Ocean, Indian Ocean, as shown on figure 5. The statistics per basin are calculated at 1 day resolution, in order to detect some limited problematic events (such as aerosol) and at 10 day and 1 month resolution, in order to detect a persistent regional default. As shown in figure 6, results do not vary significantly over the basins, except a negative bias in the Arctic Ocean, of about –0.3 to -0.5 K by night (with more noisy results by day).

Figure 5: Oceanic basins used by the UK Met Office for OSTIA and by the present validation
Figure 6a: Nighttime error as a function of time for several oceanic basins; each color corresponds to a basin and the basins are shown in figure 5; global results appear in black.

Figure 6b: Same as figure 6a but for daytime cases.
3.3 Discussion and concluding remarks

The validation results obtained on the MDB lead to the following remarks (for more details on the high latitude results see Poulter and Eastwood 2008):

- Nighttime results are quite satisfying with a standard deviation for all cases below 0.5K (table 3). The biases are negative by a few tenths of a degree in the Arctic.

- Daytime results show regional biases, mainly for temperatures over 30 °C in the Northern Indian Ocean, the Eastern Equatorial Atlantic and the Western Pacific. Biases apparent in the Arctic region are unstable. The daytime biases are partly induced by residual clouds or aerosols. Aerosol induced biases are due to several factors:
  - The SDI is calculated only from nighttime data. A simple compositing technique is used by day to provide daytime SDI value, which may then be several hours old. The daytime SDI information is thus of poorer quality than the nighttime one.
  - No correction for dust effect is attempted out of the SEVIRI disk. Indeed, for SST calculations inside the SEVIRI disk, we use the SEVIRI derived Saharan Dust Index (SDI), which has been proven efficient in correcting the SEVIRI SST (Merchant et al., 2006). Outside the SEVIRI disk, the use of the NAAPS AOD is too risky for deriving a correcting term. In addition we have used in the version 0 deliberately high limit and critical values of AOD.

- Limitations of the split window technique are another source of regional biases. This is illustrated by figure 7 (Merchant, 2008) showing the distribution of the bias of a NL type of algorithm, with coefficients adjusted by regression on the nighttime METOP MDB (i.e.; the “best” algorithm according to this MDB). The distribution of the biases for this NL algorithm is similar to that observed on figure 3 (bottom) for the NL operational algorithm used by day and is probably due to the inability of the split window method to cope with the specific atmospheric profiles of the Gulf of Guinea or the Equatorial Western Pacific. This problem has been under investigation at OSI-SAF through a series a visiting scientist studies (see Merchant et al., 2008a and b).

- The mask indicator or the confidence values are good predictors of the errors, as expected; it is recommended not to use the confidence level 2 data for quantitative use.

- In the case of METOP, calculating the error per oceanic basins as defined in figure 4 is not really efficient to show the geographical dependency of the error: the results are indeed rather similar whatever the basin, except for the Arctic Ocean.

![Figure 7](image_url): Distribution of biases obtained by applying a NL SST algorithm adjusted on the nighttime METOP MDB (from Merchant 2008).
4. COMPARISON AGAINST AATSR SST

An experimental validation of the METOP SST has been implemented, which is a comparison against nighttime AATSR SST data. The AATSR data are read in the so-called L3 files routinely produced in the framework of the Mersea experiment. Two types of L3 files have been used, global files produced by IFREMER and Atlantic files (60S - 60N - 100W - 45E) produced by CMS. METOP SST values are compared to nighttime AATSR dual view retrievals, which are corrected from skin to bulk by a constant correction of 0.2°C, the time difference between corresponding METOP and AATSR data being limited at 6 hours by night and 18 hours by day. The choice of using nighttime only AATSR data aims to avoid diurnal warming to affect the reference. A drawback of this option is that validation in Arctic in summer (respectively in Antarctic in winter) is practically impossible. Files of SST differences are produced for each METOP global file (i.e. twice daily at 0h and 12h TU) then averaged over 10 days or 1 month, separately for nighttime and daytime cases.

Figure 8: MDB mean error (top) and METOP-AATSR difference (bottom) for February 2008, decade 3, night time cases (note the color scales are not identical).
Figure 8 present the global nighttime difference of February 2008, decade # 3 (from the 21rst till the end of the month) together with the corresponding mean error map derived from the MDB. Both images are in good agreement but the METOP/AATSR difference shows better the negative error in Equatorial Atlantic and shows other features, such as a positive error in the Eastern Equatorial Pacific, which is not seen by the MDB by lack of in situ data.

Due to practical reasons, the METOP/AATSR comparison has been done for the Atlantic L3 files on the whole validation period, July 2007 to June 2008, while the global L3 files have been used only on limited periods. So the results presented in this section mainly concern the Atlantic comparison.

Figure 9 presents the night time difference METOP – AATSR over two months in order to illustrate the temporal and spatial variations of this difference.

Figure 9: nighttime METOP- AATSR difference in October 2007 (left) and February 2008 (right)

Figure 10: regions used by the METOP/AATSR comparison over Atlantic
As for the MDB validation, nighttime and daytime statistics of the difference METOP - AATSR, are calculated per zones at three resolution: 1 day, 10 day and 1 month. For the global comparison, the zones are the OSTIA oceanic basins and, for the Atlantic comparison, the smaller regions presented in figure 10. These regions have been chosen empirically, looking at monthly maps such as those presented in figure 9.

The mean and standard deviation of the SST difference for the various regions are presented in figure 11 and the results do vary from one region to another. For night time cases, a negative difference is observed in the Artic Ocean (as with the MDB); two regions, Brazil and Gulf_of_Guinea, show a negative difference, which is worst from February to May, the Red_Sea also shows negative values but highly variable with time; two regions, Galapagos and SE_Pacific, show a weak positive bias. For daytime cases, the standard deviation results are worst, as usual; the mean difference becomes positive for most of the regions, which is consistent with a possible diurnal warming; the most striking fact is that negative values over Brazil, Gulf_of_Guinea and Red_Sea are stronger due to a likely failure of the correction of the Saharan dust.

The overall statistics of the METOP – AATSR difference, over the L3_Atlantic area and the whole period, are:

- **Nighttime**
  - Mean: 0.01
  - Stddev: 0.44
  - NBC: 27698224

- **Daytime**
  - Mean: 0.05
  - Stddev: 0.66
  - NBC: 18919764

The mean and standard deviation here above are rather comparable to those of table 3, even if the METOP/AATSR comparison is completely independent from the MDB validation.

![Figure 11a](Image)

**Figure 11a**: nighttime METOP-AATSR difference as a function of time for several regions over the Atlantic; each colour corresponds to a region and the regions are shown in figure 8.
Some concluding remarks concerning the METOP/AATSR comparison:

- This comparison is a valuable tool but is not supposed to replace the MDB validation, since the reference in terms of SST remains the in situ measurements,
- This comparison gives results consistent with the MDB results, but with a better description of the regional variations (due to an improved spatial coverage),
- Regions smaller than the oceanic basins are needed in order to assess the geographical and temporal variation of the error,
- In the future, a global comparison should be done routinely, depending on the availability of the global L3 files.

Figure 11b: same as figure 11a but for daytime cases.
5. VALIDATION BY USERS

A routine validation of the OSI-SAF METOP SST products has been made by the UK met office (UKMO) and IFREMER. The DMI has produced a report on the performances of various satellite SST products, including METOP SST in July 2007.

5.1 Controls by UKMO

The UK met office have been using the METOP SST data in a test mode since spring 2007. Graphs such as those displayed on figure 12 are available on a daily basis through: http://ghrsst-pp.metoffice.com/pages/latest_analysis/sst_monitor/monitor_op/index.html

They show the results of the difference between the METOP/AVHRR SST minus the background SST, which represents a blending of various satellite data bias corrected to the AATSR, and in situ measurements. They show essentially no biases by regions, except positive biases in the Arctic and the Mediterranean. The positive bias in the Arctic, which at this time of the year is under daytime conditions, could be due to the AATSR being negatively biased in daytime, as shown in Poulter and Eastwood 2008. The positive bias in the Mediterranean Sea is probably due to OSTIA using buoy measurements there that are erroneous (and negatively biased). The problem of erroneous buoy measurements in Eastern Mediterranean has been already mentioned in section 3. Note that after several months of validation and testing, UK MO have started ingesting operationally the METOP SST granules from the 1st April 2008.

Figure 12: Regional comparison of METOP SST data with the background field before ingestion by OSTIA (from Stark, 2008); the regions are those of figure 4 with addition of the Mediterranean Sea.

Graphs such as those displayed on figure 12 are available on a daily basis through: http://ghrsst-pp.metoffice.com/pages/latest_analysis/sst_monitor/monitor_op/index.html

They show the results of the difference between the METOP/AVHRR SST minus the background SST, which represents a blending of various satellite data bias corrected to the AATSR, and in situ measurements. They show essentially no biases by regions, except positive biases in the Arctic and the Mediterranean. The positive bias in the Arctic, which at this time of the year is under daytime conditions, could be due to the AATSR being negatively biased in daytime, as shown in Poulter and Eastwood 2008. The positive bias in the Mediterranean Sea is probably due to OSTIA using buoy measurements there that are erroneous (and negatively biased). The problem of erroneous buoy measurements in Eastern Mediterranean has been already mentioned in section 3. Note that after several months of validation and testing, UK MO have started ingesting operationally the METOP SST granules from the 1st April 2008.

5.2 Controls by IFREMER

Similarly, IFREMER use the METOP data in a test mode and produce maps and graphs showing the results of their daily comparisons with the AATSR data. See: http://www.mersea.eu.org/Satellite/sst_validation_l4_glob_eur-l2p-avhrr_metop_a.html
Figure 13: Estimate of the global bias of the METOP SST compared to the AATSR for the 8th of July 2008

Figure 14: Histogram of the differences between the METOP global SST and the AATSR SST on the 8th of July 2008 before (left) and after (right) having applied the correction field deduced from figure 12.

IFREMER calculate on a daily basis the differences between the METOP global SST and the AATSR SST. A daily bias is estimated after analysis of the daily differences over the 10 days preceding the day being treated. Figure 13 shows such an estimate for the 8th of July 2008, and Figure 14 show the histograms of the differences between the METOP SST and the AATSR SST before and after the correction deduced from the bias estimate is applied.

Although no time series are available that would enable a global view of the METOP performances as evaluated by IFREMER, the 8th of July case seems rather representative of their results. Note that here again the main pattern of figure 13 is a positive bias of the METOP SST against the AATSR in Arctic.

5.3 Controls at DMI

DMI has done a detailed validation of satellite SST products over the North Sea and the Baltic in the framework of the ECOOP project. Unfortunately at the time of their work (second half of 2007) only the month of July 2007 (i.e; the very first month of METOP SST) was available.
Figure 15: Performances of METOP SST as determined by DMI over the North Sea and the Baltic (from Hoeyer, 2007), in function of the confidence levels: 3 = red, 4 = blue, 5 = green; top: biases, bottom: standard deviation

The performances of observed for the METOP SST in these conditions (figure 15) are comparable to the other OSI-SAF AVHRR products (NAR-17 and NAR-18).
6. RESULTS OF THE CMS TEST CHAIN

As noted in section 2, the version 0 chain under validation (which will be called “ope” in this section) have known limitations whose effects were clearly apparent after some months of use. To study possible improvements, a test chain (“test”) has been implemented at the end of February 2008 for CMS internal testing only. Compared to the version 0, the following modifications were tested:

1) Where there is no gradient climatology, a threshold value of 0.6 K km\(^{-1}\) on the 11 micron brightness temperature (T11) has been used. This corresponds to the maximum value normally encountered except in the very intense frontal regions of the oceans. In cases of no gradient climatology, local gradients above 0.6 K km\(^{-1}\) lead to eliminate the SST values. For local gradients below 0.6 K km\(^{-1}\) the gradient indicator is set to 50 (in place of 100 previously) which corresponds here to improving the confidence value of the cases.

2) To limit the vulnerability to residual cloudiness by day, the thresholding of T11 gradients has been also adopted by day.

3) To reduce the aerosol contamination:
   a. The SDI critical value has been restricted from 0.8 to 0.5
   b. The AOD limit and critical values have been set to 0.1 and 0.3 (against 0.5 and 1.5 previously)

4) The number of quantitatively usable data (corresponding to PCV 3, 4 and 5) has been restricted by a drift of the mask axis thresholds from [20,35,50] to [20,30,40]. In clear, PCV 2 will be attributed to cloud mask indicator values above 40 (against 50 previously), etc..

This test chain has been applied for the 3 full months of the test period: March, April and May 2008.

### Table 3: Comparison of the version 0 chain nighttime results (ope) and the CMS test chain (test) for the 3 full months of the test period, for confidence values 3 to 5 (usable) and for each confidence value

<table>
<thead>
<tr>
<th>NIGHT</th>
<th>mars</th>
<th>avril</th>
<th>mai</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>biais</td>
<td>std</td>
<td>nb cas</td>
</tr>
<tr>
<td>3 à 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>test : glo</td>
<td>0.01</td>
<td>0.43</td>
<td>12634</td>
</tr>
<tr>
<td>ope : glo</td>
<td>0</td>
<td>0.44</td>
<td>13144</td>
</tr>
<tr>
<td>test : glo</td>
<td>-0.15</td>
<td>0.64</td>
<td>359</td>
</tr>
<tr>
<td>ope : glo</td>
<td>-0.09</td>
<td>0.6</td>
<td>559</td>
</tr>
<tr>
<td>test : glo</td>
<td>0.09</td>
<td>0.56</td>
<td>2474</td>
</tr>
<tr>
<td>ope : glo</td>
<td>0.05</td>
<td>0.34</td>
<td>5186</td>
</tr>
</tbody>
</table>

### Table 4: Same as table 3, but for daytime results

<table>
<thead>
<tr>
<th>DAY</th>
<th>mars</th>
<th>avril</th>
<th>mai</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>biais</td>
<td>std</td>
<td>nb cas</td>
</tr>
<tr>
<td>3 à 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>test : glo</td>
<td>0.22</td>
<td>0.55</td>
<td>15285</td>
</tr>
<tr>
<td>ope : glo</td>
<td>0.17</td>
<td>0.58</td>
<td>17363</td>
</tr>
<tr>
<td>test : glo</td>
<td>0.06</td>
<td>0.72</td>
<td>290</td>
</tr>
<tr>
<td>ope : glo</td>
<td>0.06</td>
<td>0.72</td>
<td>291</td>
</tr>
<tr>
<td>test : glo</td>
<td>0.22</td>
<td>0.71</td>
<td>2508</td>
</tr>
<tr>
<td>ope : glo</td>
<td>0.18</td>
<td>0.74</td>
<td>2601</td>
</tr>
<tr>
<td>test : glo</td>
<td>0.17</td>
<td>0.56</td>
<td>3825</td>
</tr>
<tr>
<td>ope : glo</td>
<td>0.13</td>
<td>0.59</td>
<td>4741</td>
</tr>
<tr>
<td>test : glo</td>
<td>0.24</td>
<td>0.48</td>
<td>8722</td>
</tr>
<tr>
<td>ope : glo</td>
<td>0.2</td>
<td>0.51</td>
<td>9630</td>
</tr>
</tbody>
</table>
Table 3 and 4 report the main results of the tested chain (test) compared to the version 0 chain (ope) being validated. Figure 16 shows the distribution of the results as a function of the confidence level.

![Operational and test chain results by night](image1.png)

![Operational and test chain results by day](image2.png)

Figure 16: comparison of the results for the “ope” chain (black lines) and the test chains (red lines) in nighttime (top) and daytime (bottom) conditions, as a function of confidence levels. Biases are shown on the left and standard deviation on the right column.

This experiment led to the following remarks:

1) For the quantitatively usable data (confidence values 3 to 5), the test chain shows significantly better statistics than the ope chain.
2) The improvement concerns all classes 3 to 5 by day and particularly class 3 and 4 by night.
3) The performances for class 2 for the test chain are degraded (by night) which is a consequence of transferring more dubious cases to this category.
4) The number of cases in all classes have diminished (this results from more severe cloud control), except in class 5 by night, benefiting likely of the increase of confidence brought by the modification of the indicator value in case of missing gradient climatology.
5) The detailed comparisons by zone (not displayed here) show that the gain in performances has been mainly achieved over the Northern Atlantic, the Indian and the Southern ocean (see zone definition in figure 5), an indication that indeed residual cloudiness and aerosols have been better accounted for.

Since the results were rather stable throughout the period and corresponding to our expectations, the test chain has been stopped in June 2008.
7. CONCLUSION

7.1 Summary of the results

The global statistics are well within the target accuracy defined in the OSI SAF CDOP requirement documents (monthly biases and standard deviation < 0.5K and 0.8K, respectively). The bias and standard deviation calculated over 11 months (July 2007-June 2008) are indeed ~0.01K and 0.46K respectively for nighttime cases, and 0.12K and 0.58K for the daytime cases. They are fairly stable with time by night and show a small seasonal cycle by day. The regional distribution of the errors shows no convincing differences between the basins as defined by UKMO and adopted by GHRSST, except for the Arctic, where a deficit has been observed. According to Poulter and Eastwood, this deficit is about ~0.34 K by night and null by day once the twilight cloud mask issue is corrected for. They have also identified a positive bias of about the same amplitude by day in the Antarctic. The mapping of the errors, as well as the comparisons with the AATSR have shown negative biases in specific areas of the oceans such as the Eastern Equatorial Atlantic and the Western Equatorial Pacific. These biases are observed for most IR sensors, including SEVIRI and the OSI-SAF have launched several studies to try to solve this (see Merchant et al 2008a and 2008b), but short term solutions in the frame of the multispectral approach seem unlikely. Although the origin of these biases is still not completely clear, they demonstrate the limitations of the classical multispectral approach that has been, so far, universally adopted in operational SST retrieval from IR imagery.

The validation/controls done by users have confirmed the main characteristics of our own results. UKMO have started ingesting operationally the METOP SST granules in their operational analysis from the 1st April 2008.

7.2 Modifications to the chain

Lessons drawn from this validation experiment have led to the definition of a new processing version (version 1). In a first step a test version was implemented internally at CMS in February 2008 to test the modifications that were inspired by the first months of validation results. These modifications were centered on more strict cloud edges elimination by reinforcing the use of local T11 gradients, and lowering the thresholds on the aerosol information to better eliminate the potentially contaminated cases. After the conclusions of Poulter and Eastwood, the use of the met.no ice/cloud mask has been extended to the twilight zone (for values of the sun zenith angle between 85 and 95 degrees). This set of changes has been implemented on the 24 June 2008 as the “version1” processing chain, in place of version 0. The results expected from the version 1 with respect to version 0 is a reduction of the cloud/aerosol residual contamination, hence a reduction of the error standard deviation, and the elimination of the twilight cloud mask problems. This version 1 is candidate for being “operational”. No modification of the algorithms are envisaged for the moment, since no immediate solutions have been found for the two issues that could be algorithm related, i.e the cold biases at night in the Arctic, and for temperatures over 30C. However we will continue investigating the problem in relation with met.no, along the lines suggested in Poulter and Eastwood, and the University of Edinburgh.

7.3 Related issues

Aside the METOP SST validation results, this validation experiment has revealed the following issues, that will likely be discussed within the SST community (the GHRSST group):

1) Validation Basin definition: the results observed with METOP have shown the large basins adopted by the Met Office not to be adapted for tracking the METOP regional errors which appear at a smaller scale. Comparison with AATSR at CMS may lead to proposing a new partition of the Ocean. However, the trade off between a precise evaluation of the regional errors and a too large number of validation zones may not be trivial to establish.

2) Erroneous buoy measurements: operational quality control (QC) procedures have been established for long by the Joint Commission for Oceanography and Marine Meteorology in situ Observing Platform Support Center (JCOMMOPS). The SST METOP validation
experience has shown that satellite in situ measurements intercomparison may reveal errors in buoy measurements in the range of 1 to 2K, which are hardly detectable otherwise. The GHRSST-PP meeting in Perros-Guirec (June 2008) has been the opportunity to open a fruitful dialogue with the JCOMMOPS people in charge of the operational buoy measurement quality control, who envisage using now the METOP MDB to identify erroneous buoy data.

References


Hoyer, J. (2007), Validation of high resolution GHRSST-PP satellite sea surface temperature observations in the North Sea/Baltic Sea region, ECOOP WPO3-01.03.


