Deep-Sea Research II 77-80 (2012) 21–30

Contents lists available at SciVerse ScienceDirect

Deep-Sea Research II

journal homepage: www.elsevier.com/locate/dsr2

Group for High Resolution Sea Surface temperature (GHRSST) analysis fields inter-comparisons. Part 1: A GHRSST multi-product ensemble (GMPE)

Matthew Martin[a,*], Prasanjit Dash[b,c], Alexander Ignatov[b], Viva Banzon[d], Helen Beggs[e], Bruce Brasnett[f], Jean-Francois Cayula[g], James Cummings[h], Craig Donlon[i], Chelle Gentemann[j], Robert Grumbine[k], Shiro Ishizaki[l], Eileen Maturi[b], Richard W. Reynolds[m], Jonah Roberts-Jones[a]

[a] Met Office, Exeter, UK
[b] NOAA/NESDIS, Center for Satellite Application and Research (STAR), Camp Springs, MD, USA
[c] CSU, Cooperative Institute for Research in the Atmospheres (CIRA), Fort Collins, CO, USA
[d] NOAA/NESDIS, National Climatic Data Center (NCDC), Asheville, NC, USA
[e] CAWCR, Bureau of Meteorology, Melbourne, Australia
[f] EC CMC, Dorval, Quebec, Canada
[g] QinetiQ North America, Technology Solutions Group, Stennis Space Center, MS, USA
[h] Naval Research Laboratory, Monterey, CA, USA
[i] ESA/ESTEC, Keplerlaan 1, 2201 AZ, Noordwijk, The Netherlands
[j] Remote Sensing Systems, Santa Rosa, CA, USA
[k] NOAA/NCEP, Camp Springs, MD, USA
[l] JMA, Tokyo, Japan
[m] NOAA/CICS, Asheville, NC, USA

ARTICLE INFO

Available online 2 May 2012

Keywords:
Sea surface temperature
Intercomparison
Ensemble

ABSTRACT

Many sea surface temperature (SST) gap-free gridded analysis (Level 4, or L4) fields are produced by various groups in different countries. The Group for High Resolution SST (GHRSST) is an international collaboration body which has formed the inter-comparison technical advisory group (IC-TAG), to advise SST producers and users on the relative performance of these SST fields. This two-part paper describes two of the three major systems developed under GHRSST coordination towards this goal. Part one (this paper) describes the GHRSST Multi-Product Ensemble (GMPE) system, which runs on a daily basis at the UK Met Office, taking various L4 analyses as inputs, transferring them onto a common grid, and producing an ensemble median and standard deviation. The various analysis systems contributing to the GHRSST inter-comparisons are discussed, highlighting areas of commonality between the systems as well as those parts of the systems where there is less agreement on the appropriate algorithmic or parametric choices. The characteristics of the contributing L4 analyses are demonstrated by comparing them to near-surface Argo profile temperature data, which provide an independent measurement of SST and have been shown to provide a good estimate of foundation SST (the SST free of diurnal warming). The feature resolution characteristics of the L4 analyses are demonstrated by calculating horizontal gradients of the SST fields (on their original grid). The accuracy and resolution of the GMPE median are compared with those of the input analyses using the same metrics, showing that the GMPE median is more accurate than any of the contributing analyses with a standard deviation error of 0.40 K globally with respect to near-surface Argo data. For use in climate applications such as trend analysis or assimilation into climate models, it is important to have a good measure of uncertainty, so the suitability of the GMPE standard deviation as a measure of uncertainty is explored. This assessment shows that, over large spatial and temporal scales, the spread in the ensemble does have a strong relationship with the error in the median, although it underestimates the error by about one third.

1. Introduction

Sea surface temperature (SST) is an important variable for many applications covering a large range of space and time scales. Operational uses of SST data include seasonal forecasting in which SST information is used to constrain the air–sea interface during the analysis and reanalysis phases, climate monitoring, numerical weather prediction (NWP) in which SST information is used to prescribe the lower temperature boundary condition over the ocean, and short-range ocean forecasting in which SST data are assimilated during initialization (Le Traon, 2011) and used for

* Correspondence to: Met Office, Fitzroy Road, Exeter, UK. Tel.: +44 1392 886465.
E-mail address: matthew.martin@metoffice.gov.uk (M. Martin).

0967-0645/$ - see front matter Crown Copyright © 2012 Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.dsr2.2012.04.013

validation (Martin, 2011). A large number of research users also exist, including those using SST data for climate change studies, and SST retrieval groups which often require gap-free SST fields to provide a first guess in radiative transfer simulations (e.g., Merchant et al., 2008, 2009; Liang et al., 2009; Le Borgne et al., 2011; Petrenko et al., 2011). Dash et al. (2010) and Stark et al. (2008) have also used global SST fields for self- and cross-consistency checks of various lower level SST data products, and found them a useful supplement to the customary validation against in situ SST data.

SST data have been collected using in situ measuring techniques for over 150 years (Rayner et al., 2003). Satellite measurements of SST have been available since the 1970s, with a continuous record of accurate SST data from the Advanced Very High Resolution Radiometer second generation (AVHRR/2) instruments on board the NOAA polar orbiting satellites available since 1981, followed by the third generation (AVHRR/3) onboard NOAA satellites after the launch of NOAA-15 in 1998 and the MetOp-A satellite, the latter operated by EUMETSAT since 2006. Other infrared instruments on board polar orbiting satellites include the Along-Track Scanning Radiometer (ATSR) series of instruments on board ESA satellites since 1991, and the Moderate Resolution Imaging Spectroradiometer (MODIS) instruments on board the Terra and Aqua satellites operated by NASA since 2000 and 2002, respectively. Geostationary satellites also provide information about SST, for instance from the GOES (NOAA), Meteosat (EUMETSAT) and MTSAT (JMA) satellite series. Microwave measurements of SST have been available since 1978 from NASA's Scanning Multichannel Microwave Radiometer (SMMR) satellite, followed in 1997 by the TRMM Microwave Imager (TMI, operated by NASA and JAXA) and the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) instrument developed by JAXA and flown on board NASA's Aqua satellite since 2002 (AMSR-E data are no longer available since 5th October 2011).

The international Group for High Resolution SST (GHRSST[1]) was set up to facilitate the exchange of satellite SST information and to meet the needs of the varied users. GHRSST facilitates the coordination and exchange of ideas for developing the various levels of data processing, from the radiances measured by the satellite instruments, through the SST retrieval process producing Level 2 data (SSTs at the observed pixels), Level 3 data (products gridded in space, but with no gap-filling), to the production of gap-free objective analyses called Level 4 (or L4) data. A large proportion of the applications described above require the SST data as global, gap-free analyses, and the focus of this paper is on these L4 products.

The techniques used to combine the SST information available from the various in situ and satellite platforms into high resolution global gap-free analyses have been developing over the past decade. A large number of these L4 products now exist, each with its own characteristics. To facilitate their cross-evaluation and optimal use, three L4 monitoring and inter-comparison systems have been set up under the GHRSST umbrella: the GHRSST Multi-Product Ensemble (GMPE) which provides near-real time (NRT) global ensemble information from a number of input L4 products; the L4 SST Quality Monitor (L4-SQUAM) which monitors the quality of the various L4 products globally in NRT; and the High-Resolution Diagnostic Data Set (HR-DDS) which inter-compares various SST data-sets (satellite-based and in situ) at predefined locations. The L4-SQUAM system is aimed at cross-comparisons of the various L4 analyses using a number ways of presenting statistics, whereas GMPE is aimed at generating an ensemble estimate, and estimating deviations of various L4 analyses from it. The development of these systems and their complementary functionality is coordinated through the Inter-Comparison Technical Advisory Group (IC-TAG) within GHRSST.[2] The GMPE system is described here, the SQUAM system is described in the part 2 companion paper (Dash et al., this issue) and the HR-DDS system is documented by Donlon et al. (2009).

The GMPE system is relevant to a number of applications of SST for climate. For instance, the monitoring of the current state of the climate requires robust information about SST close to real time. This information can be provided from individual analyses, but a consensus analysis (as might be produced from an ensemble product like GMPE) has the potential to be more accurate, robust and stable than any individual analysis. The possibility of providing information about uncertainty in the ensemble analysis using the spread in the ensemble is also potentially very useful for climate applications where accurate uncertainty measures are a key requirement. There are plans to produce GMPE back to 1991 using the available SST reanalysis products which will have a number of uses for climate applications.

The main aims of part one are to describe the GMPE system, with a particular emphasis on an inter-comparison of the various L4 analyses to enable users to choose those most suitable for their respective applications. This work complements previous comparisons of SST fields such as that of Reynolds and Chelton (2010). Section 2 describes various L4 products contributing to GHRSST and Section 3 describes the method used to combine some of these L4 products into the GMPE with examples of the use of GMPE. Section 4 describes an assessment of the accuracy and bias in the various L4 analyses and the GMPE median using independent Argo data, provides some examples of the horizontal gradients in each of the L4 analyses and the GMPE median, and assesses the potential for using the GMPE standard deviation as a proxy for error in the median value. Section 5 summarizes the results and suggests some areas for future research and development.

2. L4 products contributing to GMPE

This section describes the characteristics of each L4 system developed in Australia, Canada, Japan, the UK, and the USA that contribute to the GMPE. It is not possible to detail all the aspects of the systems here, rather we focus on those aspects of relevance to the results described in Section 4 of this part one and in part two. Most of the systems described here provide their data free of charge through the GHRSST Global Data Assembly Center (GDAC[3]) in a common format described by the GHRSST Data Specification. The files contain the SST analysis itself, an error estimate, and information about sea-ice concentration.

An overview of most of the global L4 analysis systems currently contributing to the GHRSST is shown in Table 1. All of these systems produce SST analyses on at least a daily update cycle in near-real time (NRT) with at least 0.25° horizontal resolution using a combination of satellite SST data-sets, with some of the systems also including in situ SST data.

Most of these systems have carried out reanalyses as well as NRT production, although these reanalyses are of differing lengths. The OISST.v2 AVHRR and MGDSST products have the longest record, starting from September 1981, with OSTIA having reanalyses available from January 1985. These reanalyses generally make use of re-processed versions of the various data-sets which are not available in near-real time. However, an

[1] http://www.ghrsst.org.

[2] https://www.ghrsst.org/ghrsst-science/science-team-groups/ic-tag.

[3] http://ghrsst.jpl.nasa.gov.

Table 1
Summary of the various L4 SST analysis systems.

Name of system	*Institute, country*	*Reference*	*Date of availability of NRT data*	*Reanalysis period*
CMC	Canadian Meteorological Center, Canada	Brasnett (2008), Brasnett (1997)	April 2008	September 1991
FNMOC	Fleet Numerical Meteorology and Oceanography Center, USA	Cummings (2005), Cummings (2011)	2005	N/A
GAMSSA	Bureau of Meteorology, Australia	Beggs et al. (2011), Zhong and Beggs (2008)	October 2008	N/A
MGDSST	Japan Meteorological Agency, Japan	Kurihara et al. (2006)	September 2002	September 1981
NAVO K10	Naval Oceanographic Office, USA	–	July 2004	N/A
OISST.v2:AVHRR, AVHRR+AMSR	NCDC/NOAA, USA	Reynolds et al. (2007)	November 2008 (V1 started 2006 and was discontinued in 2009)	September 1981 (AVHRR), June 2002 (AVHRR+AMSR)
OSTIA	Met Office, UK	Donlon et al. (2011)	November 2006	January 1985
POES/GOES	NESDIS/NOAA, USA	Maturi et al. (2008)	June 2007	N/A
RSS MW, MW/IR	Remote Sensing Systems, USA	Gentemann et al. (2006)	2007	June 2002, January 2002
RTG	NWS/NCEP/NOAA, USA	Gemmill et al. (2007)	September 2005	N/A

Table 2
Summary of the data types used by the various L4 SST analyses in NRT. Note that AMSR-E data are not available since 5th October 2011.

Name of system	*In situ*	*AATSR/Envisat*	*AVHRR/NOAA*	*AVHRR/MetOp*	*MODIS/Aqua,Terra*	*AMSR-E/Aqua*	*TMI/TRMM*	*SEVIRI/MSG*	*GOES*	*MTSAT-2*
CMC	X	X	X	X		X				
FNMOC	X		X	X					X	
GAMSSA	X	X	X	X		X				
MGDSST	X		X	X		X				
NAVO K10			X	X		X			X	
OSTIA	X	X	X	X		X	X	X		
POES/GOES		X	X	X		X		X	X	X
OISST.v2:AVHRR	X		X	X						
OISST.v2:AVHRR+AMSR	X		X	X		X				
RSS-MW						X	X			
RSS-MW/IR					X	X	X			
RTG	X		X	X						

assessment of the NRT products provides a good indication of the characteristics of the delayed-mode versions of each of the systems.

2.1. Input data types

The data types used in the production of the various NRT L4 analyses are summarized in Table 2. The most commonly used data-sets are the in situ data, the AVHRR data from NOAA and MetOp, and the AMSR-E data (the results in later sections of this paper are for the year 2010 when AMSR-E data were available—these are no longer available since 5th October 2011, although other microwave data are available and are being investigated to replace AMSR-E data in some of the GHRSST systems). These data-sets have complementary features: the AVHRR data are reasonably accurate with global high resolution coverage, the AMSR-E data are less accurate but provide information even in the presence of clouds, the in situ data can provide a means of correcting for biases in the various satellite data. The geostationary data from GOES, SEVIRI and MTSAT-2 also provide complementary information over specific regions, as they have much higher temporal resolution, although these are currently less-widely used by the various analysis systems. The in situ data used by most of the systems are from moored and drifting buoys, and ship-mounted sensors.

2.2. Quality control (QC) and input data processing

Quality control of input data is a crucial step in the processing chain for the operational L4 systems. Allowing bad data into an analysis can have an adverse impact on the result, whilst a too-stringent QC procedure can mask observations of anomalous events which are crucial for users. For example, a minimum in the sea-ice extent in the Arctic summer in 2007 was associated with a large positive SST anomaly which was up to 10 K warmer than the climatological average. This led a number of the quality control systems to reject the observations in that region (Donlon et al., 2012).

The level-2 pre-processed (L2p) satellite data provided through GHRSST (the content of which is described in Donlon et al., 2007) contain information about the quality of the data, estimated by the data providers. These QC flags are used by most of the L4 systems while some (e.g., OISST.v2) use the best quality level 3 data. On top of these checks, most of the systems perform their own QC. These include statistical checks against the previous analysis (or the current background field) and/or against climatology. Additional checks are performed by some systems. For instance, the RSS system also screens the data based on the cloud screening of MODIS, and rain screening of AMSR-E and TMI. For the CMC, FNMOC and GAMSSA systems, checks are also included whereby data within a prescribed distance of each other are compared, and any outliers are removed.

Most of the systems also perform a check to prevent contamination by diurnal warming, since they aim to provide an estimate of foundation SST. This is usually done by rejecting day-time data if the wind speed is less than 6 m/s. Some systems (e.g., GAMSSA) also reject night-time data in low wind-speed conditions. The MGDSST system rejects satellite data if the diurnal amplitude is larger than 3 K. The RSS MW and MW/IR analyses both use satellite SST data which have been adjusted to a foundation SST using an empirical diurnal model (Gentemann et al., 2003). The FNMOC system uses a 6-h update cycle and therefore includes a diurnal cycle, although diurnal warming events are flagged (where the anomaly from the previous analysis is significantly warmer than expected, the wind speed is less than 6 m/s and the solar radiation exceeds 100 W/m^2). The flagged data are used in the 6-h update cycle analysis for GHRSST but are not used in the 24-h update cycle used for NWP applications.

In situ data checks are performed by some of the systems. These include statistical background checks and/or climatological checks, with some systems (e.g., CMC and FNMOC) also performing checks to eliminate reports likely to have errors in position, reports likely to be biased and platforms with erratic previous reports.

The raw L2p data provided through GHRSST are at very high resolution, leading to extremely large numbers of data available each day. For instance, the MetOp AVHRR data is available at 1-km resolution globally. Geostationary L2p data files include SEVIRI, GOES, MTSAT-2 every 15 min, 30 min, and hourly, respectively. Since most of the L4 analyses are calculated at lower spatial and temporal resolution, some thinning/averaging of the data is performed by most systems. This is usually done using a median filter to avoid using outliers. Some of the systems average the data from a particular satellite within analysis grid cells (e.g., GAMSSA), and the FNMOC system also averages within particular water mass classifications (defined using a Bayesian water mass classification technique) which helps maintain cross-frontal SST gradients in the analysis.

2.3. Bias correction

The GHRSST L2p files contain information about the expected systematic and random components of the error for each observation (called Sensor Specific Error Statistics, or SSES). The SSES bias values are subtracted from the observations before use in the objective analysis by most of the L4 systems. However, time-varying biases in the observations still exist, and the various observation types can be biased compared to one another due to the varying measurement techniques and retrieval algorithms used. A number of L4 systems therefore have an additional bias correction step in order to remove these biases. The techniques for doing this vary between the systems but generally involve the use of a reference data-set (usually a sub-set of the in situ data, sometimes including the AATSR data) which is used to correct each satellite data type in turn (see for example Donlon et al., 2012).

2.4. Background field

For most of the L4 systems, the background field (used as a basis for the objective analysis described in the next sub-section), is based on the analysis of the previous day with a relaxation to climatology. In ice-free regions, the e-folding time-scale for this relaxation varies from 58 days (CMC) to about 10 days (OSTIA), although some of the systems retain the full analysis from the previous cycle as the background (e.g., OISST.v2, FNMOC, NAVO K10).

Some L4 systems use a relaxation scheme to provide an SST estimate under sea-ice that is consistent with sea ice concentration. Under sea-ice, SSTs are relaxed towards freezing with a shorter time-scale than for ice-free regions. For example, in OSTIA the SSTs under sea-ice with concentration $>50\%$ are relaxed towards 271.35 K with an e-folding time of 17.5 days at 50% which decreases to 5 days at 100% ice concentration. Other systems deal with sea-ice/SST consistency by other methods. For instance, the FNMOC system analyzes sea-ice concentration and SST simultaneously using an observation operator which generates freezing SST observations when the sea-ice concentration exceeds 55% (using a climatological salinity field to vary the freezing temperature with location).

2.5. Objective analysis

All of the analysis systems described here, with the exception of NAVO K10, solve the objective analysis equation:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{B}\mathbf{H}^T[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}]^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}^b) \quad (1)$$

where the analysis $\mathbf{x}^a$ (containing values x_i^a, $i = 1,\ldots,N$ at all points on the analysis grid) is obtained by adding onto a background field (the generation of which is described in the previous section), $\mathbf{x}^b$, a weighted difference between the observation vector, $\mathbf{y}$, and the background field interpolated to the observation locations using an observation operator $\mathbf{H}$. The weighting is determined by specifying the error covariance matrices for the background, $\mathbf{B}$, and observations, $\mathbf{R}$. The analysis systems solve Eq. (1) using the Optimal Interpolation method or variants of it. The details of how these are implemented in the various systems are described in more depth in each of the references in Table 1.

The specification of error covariance matrices is a crucial aspect of each of the systems, particularly since there is no model to evolve the information as might be used in a data assimilative system (other than the relaxation to climatology). The observation error covariance matrix is assumed in most of the systems to be a diagonal matrix, i.e. observations are uncorrelated, although the GAMSSA system includes the effect of correlated observation errors. The observation error variances are generally specified using the SSES random errors provided in the GHRSST L2p files, although some systems (e.g., CMC, FNMOC and GAMSSA) derive their own estimates of the observation error variances. The FNMOC system models the errors of representativity by taking into account the size of the satellite footprint, the resolution of the grid and the background SST gradients.

The background error covariance matrix determines the manner in which the observational information is spatially propagated in the analysis. The full matrix is very large, especially for the high resolution analyses described here, and cannot be specified completely. It is therefore parameterised by specifying a set of variances (for the diagonal elements) and spatial de-correlation length-scales which provide the parameters for a function which approximates the off-diagonal elements.

The observation operator is often assumed to be a simple bilinear interpolation from the analysis grid to the observation locations, although some systems use more complicated methods (e.g. the FNMOC method fits a surface using the surrounding 4×4 points and evaluates the surface at the observation location). For point observations, or satellite observations which are representative of a small area compared to the size of the analysis grid size, this assumption is valid. For those satellites where the observation footprint is larger than the analysis grid size, other observation operators are required. Here the background gridded values which fall within the observation footprint should be used to estimate the background equivalent of the observation.

2.6. Horizontal and temporal resolution

The resolution of the features resolved by each analysis depends on a number of factors including the horizontal grid, the input data resolution and type, and the horizontal correlation scales used in the objective analysis. The resolution of the horizontal grid, on which various products are calculated, varies between 0.25° and 0.05° as shown in Table 3.

2.7. Analysis error

The various L4 analyses currently produce analysis error estimates using different techniques. Some of them are based on the formal analysis error estimate coming out of the Optimal

Table 3
Summary of the characteristics contributing to the horizontal and temporal scales resolved by each L4 analysis.

Name of system	*Horizontal grid resolution*	*Minimum horizontal correlation scale*	*Highest input data resolution (after thinning)*	*Temporal correlation scale*	*Update cycle*
CMC	0.2° lat/lon	43 km	44 km	None	Daily by 09:15 UTC for analysis of previous day
FNMOC	~9 km	~10 km	~9 km	None	6 hourly with 7 h delay on each analysis
GAMSSA	0.25° lat/lon	50 km	17 km	0.5 days	Daily by 03:30 UTC for analysis of previous day
MGDSST	0.25° lat/lon	50 km	25 km	5 days	Daily by 00:30 UTC for analysis of previous day
NAVO K10	10 km	10 km	2 km	4 days	Daily by 03:00 UTC for analysis of previous day
OISST.v2:AVHRR, AVHRR+AMSR	0.25° lat/lon	50 km	25 km	3 days	Daily by 12:30 UTC for analysis of previous day
OSTIA	0.05° lat/lon	10 km	~3.5 km	5 days	Daily by 06:30 UTC for analysis of previous day
POES/GOES	0.1° lat/lon	11 km	4 km	None	Daily
RSS MW, MW/IR	25 km, 9.76 km	1°, 1.5°	25 km, ~10 km	4 days, 3 days	Daily
RTG	1/12° lat/lon	50 km	8 km	None	Daily by 21:30 UTC for analysis of day

Interpolation equations, whilst others are based on more ad hoc techniques. There are pros and cons to each approach since the formal analysis error estimate relies on the input information (e.g., the error covariance matrices, which are often poorly known) to be specified accurately in order to provide a good estimate of error, and other techniques have less-sound theoretical bases. An on-going focus for the GHRSST IC-TAG is to improve and attempt to standardize the way in which the L4 producers estimate their analysis error.

3. The GHRSST multi-product ensemble

The GHRSST Multi-Product Ensemble (GMPE) is produced routinely at the UK Met Office. It runs on a daily basis, taking in various L4 GHRSST products, and producing ensemble median and standard deviations over the global oceans. As of writing, the following products are included in GMPE: CMC, FNMOC, GAMSSA, MGDSST, NAVO K10, OISST.v2 AVHRR, OISST.v2 AVHRR+AMSR, OSTIA, RSS MW/IR, RSS MW, RTG.

The OISST.v2 AVHRR+AMSR product is not currently included in the real-time GMPE system but is available for the period of the study and is therefore included in the results. The FNMOC analysis is available 4 times a day but only the file valid at 12:00 UTC is used in GMPE. This will result in biases in the results shown in the next section for the FNMOC analysis since the FNMOC 12:00 UTC field will always be daytime at zero longitude and a mixture of daytime and night-time in the western and eastern Pacific for example. The ODYSSEA analysis, produced at Ifremer, France, was originally included in the GMPE system but was not available during 2010 which is the period used to generate the results shown in subsequent sections.

Each of the products in the above list is interpolated onto the same regular 1/4° latitude/longitude grid, and the median and standard deviation of their SSTs are calculated at each grid-point. Median was preferred over the conventional mean estimate, to minimize the effect of possible outliers on the ensemble estimates. The output files are freely available through the MyOcean project[4] and the format is described in the GHRSST Data Specification document (GHRSST Science Team, 2010). The main purposes for producing GMPE are to monitor the current state of the global SST, as well as each of the input L4 analyses.

GMPE is currently available going back to September 2009. However it is planned to produce a version which goes back to at least 1991 in order to provide a product which can be used for assessing the various SST reanalysis products, and to provide a new consensus estimate of global SSTs over that period. Thus an assessment of the GMPE SSTs produced in NRT is useful in the context of climate studies because it will inform the development of the longer-term GMPE, as well as providing an assessment of whether the GMPE median is a good estimate of the current state of the ocean surface temperature.

4. Inter-comparisons

This section provides examples of inter-comparisons which have been carried out using the various contributing L4 analyses to the GMPE, and the GMPE median. The main aims are to assess the accuracy of the GMPE median compared to the accuracies of each of the individual L4 analyses, to assess other aspects of the usefulness of these analyses including the horizontal resolution of features within each analysis, and to assess the possibility of using the standard deviation in the GMPE as a measure of uncertainty in the GMPE median estimate.

4.1. Validation against Argo data

Near-surface Argo data have been shown to provide a good estimate of foundation SST using a triple-collocation of Argo data with surface drifters and AATSR data (Merchant and Corlett, pers. comm., 2010). None of the contributing L4 analyses to the GMPE use Argo as a data-source, so Argo data provide the basis for an independent assessment of the accuracy of the SST products.

Argo data have been taken from the EN3 data-set.[5] These data have been taken from the Argo Global Data Assembly Center, and have undergone an additional, consistent QC using the procedures described in Ingleby and Huddleston (2007). All Argo data available in 2010, valid between 3-m and 5-m depth which passed the QC have been used here.

[4] By emailing servicedesk@myocean.eu.org.

[5] http://hadobs.metoffice.com/en3.

The GMPE median has been interpolated bi-linearly to the Argo data locations valid on each day, and mean and root-mean-squared (RMS) differences have been calculated in $2^\circ \times 2^\circ$ boxes, as shown in Fig. 1. The coverage of the Argo data is very limited in the South Pacific, Intra-Americas Seas and Indonesian regions. Generally the GMPE median has a small bias compared to the Argo data, with a mean difference of 0.03 K globally over the whole year. The RMS differences are largest in regions of high mesoscale ocean variability such as the Gulf Stream, Kuroshio, Agulhas retroflection, and Zapiola Rise region off the coast of Argentina.

Each of the input L4 analyses for a particular day has been interpolated bi-linearly from its original resolution to the Argo data locations valid on the same day as the analysis (the only exception is the CMC analysis which was provided at a slightly degraded resolution during this period—$1/3^\circ$ rather than 0.2°—which results in a slight increase in its errors of about 2–3%). The monthly mean and standard deviation of these match-ups have been calculated globally and over various ocean regions for the year 2010 for each of the analyses. These are shown in Fig. 2 for the global ocean, and for three example sub-regions (North Atlantic, Tropical Pacific, and North Pacific).

Globally the GMPE median is more accurate than any of the contributing L4 analyses with an overall standard deviation of 0.40 K. All of the L4 products have a standard deviation of less than 0.7 K with the CMC, GAMSSA, NAVO K10, and OSTIA L4 products having a standard deviation of less than 0.5 K. The overall bias in the L4 analyses, including the GMPE median, is generally in the region 0.02–0.04 K, except for the NAVO K10, FNMOC, RTG and RSS TMI/AMSR-E products which have larger biases.

In all sub-regions, the GMPE median has smaller standard deviation error than any of the input L4 analyses. The bias of the median is always less than 0.07 K, this maximum bias occurring in the North Pacific region. The North Atlantic region has the highest standard deviation error in all of the L4 analyses, most likely due to the highly variable nature of the Gulf Stream and North Atlantic current, and the presence of large amounts of cloud over these regions.

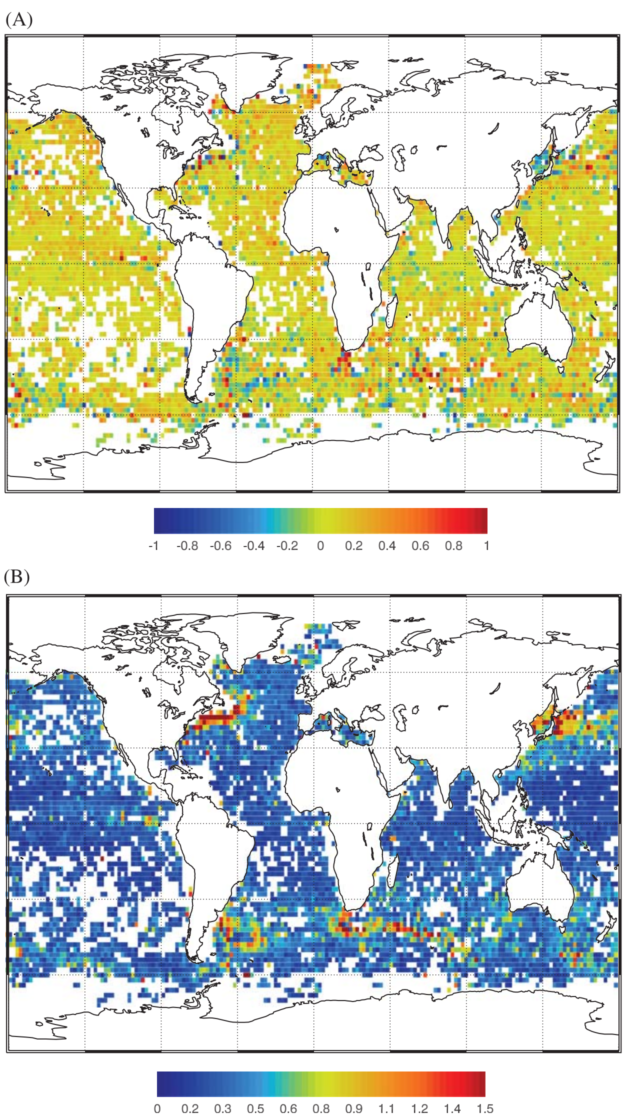

Fig. 1. (A) Mean (Argo-minus-GMPE) and (B) root mean square differences (K), in $2^\circ \times 2^\circ$ boxes, between near-surface Argo data and the GMPE median for the year 2010.

4.2. Horizontal gradients

Horizontal SST gradients are important for a number of applications since they affect aspects of atmospheric forecasts (Maloney and Chelton, 2006), and are important for localized studies of ocean circulation and climate. In order to assess the resolution of features within each of the L4 analyses and to compare the resolution of the GMPE median with those analyses, their horizontal gradients have been calculated for the 1st March 2010 as a representative example. It is difficult to provide a "truth" with which to compare these gradients as the methods used to produce gradient information directly from the level 2 satellite SST data are not perfect either (see Ullman and Cornillon, 2000). Rather than a quantitative estimate of the validity of these gradients we provide a qualitative summary of the way in which each analysis represents the gradients.

Fig. 3 shows the grid-point averaged east–west and north–south gradients of SST for each analysis in the Gulf Stream region for the 1st March 2010. These gradients have been calculated on the raw grid of each analysis and interpolated onto the same $1/4^\circ$ grid before plotting. It is clear from these figures that the different analyses provide very different information about gradients. Most of the analyses capture the main features expected in this region with its numerous mesoscale features. However, the gradients are highlighted more in some L4 products than in others. For example, they are more prominent in NAVO K10 (Fig. 3H) as compared to CMC (Fig. 3A) and OSTIA (Fig. 3E), probably due to the different set-ups of these analyses: the NAVO K10 system uses direct insertion whilst the others solve the OI anlaysis Eq. (1). Given that OI is a low-pass smoothing operation perhaps this explains the more prominent gradients in NAVO K10 than other analysis products, although it should be noted that these are not necessarily more realistic. These results highlight the difficulties in setting up the analysis parameters such as the error covariances in the sense that the aim is to produce the most accurate analysis while retaining the highest resolution features possible. If high resolution input data were available everywhere, this would be an easily attainable goal, but the inhomogeneity of the input data makes it difficult in practice to obtain high resolution features where the data is high resolution whilst not introducing spurious features where the data does not support it (although the use of adaptive error covariances such as is used in the FNMOC analysis should reduce this problem).

The ability of the GMPE median to represent the high gradient features is surprisingly good. One might expect the use of

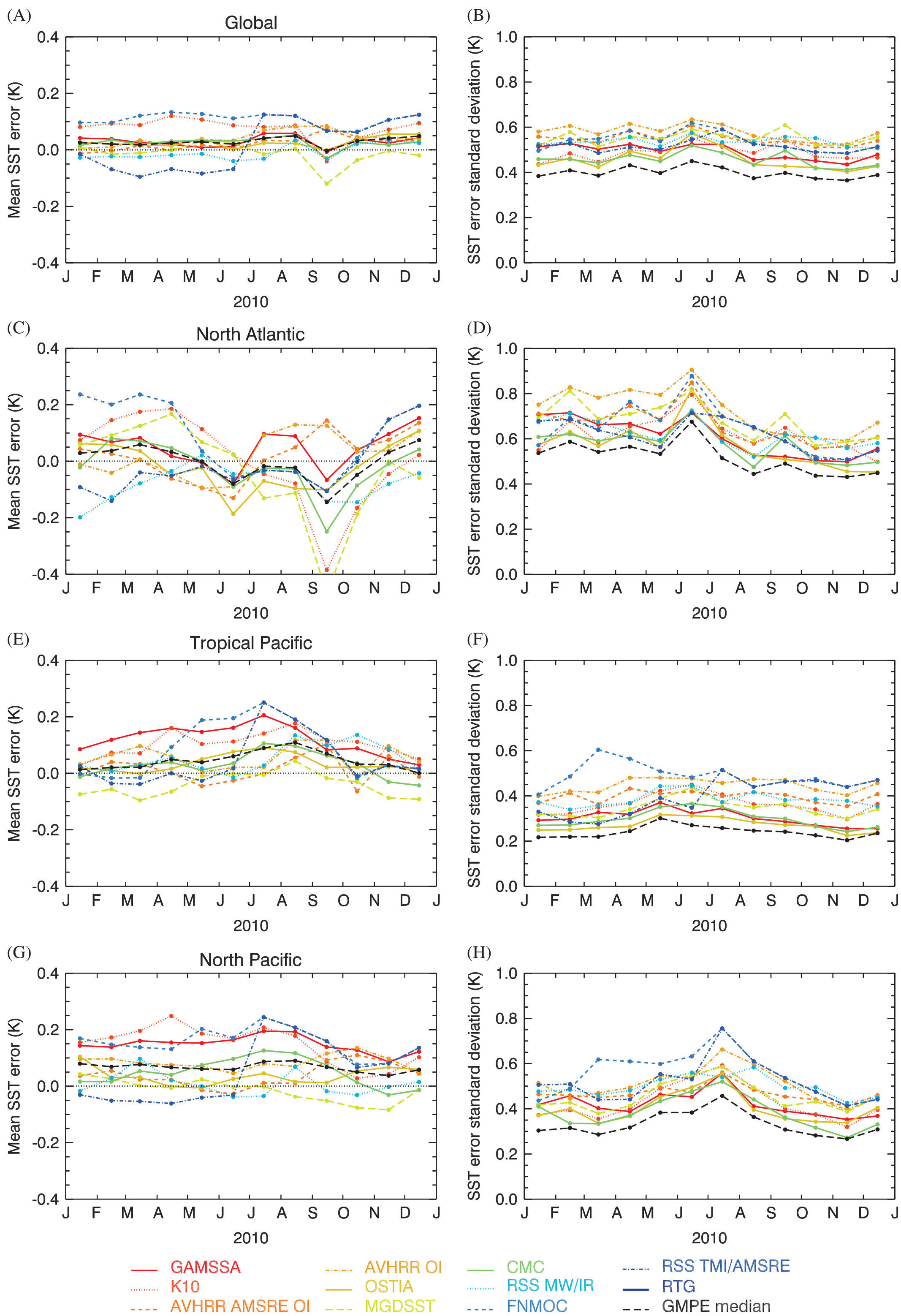

Fig. 2. Mean (Argo-minus-GMPE, left) and standard deviation (right) differences between each L4 analysis (including the GMPE median) and near-surface Argo data calculated monthly during 2010. The statistics are shown for four regions: Globally (top), the North Atlantic (second row), the Tropical Pacific (third row), and the North Pacific (bottom). A legend showing which line corresponds with each analysis is shown.

ensemble median to introduce artificial gradients when the product which is the median changes in the horizontal. In order to demonstrate how much the origin of the median changes, Fig. 4 shows the SST values of each input analysis, the GMPE median SST, and a representation of which input analysis is the median value at each grid point along 40°N between 70°W and 55°W for

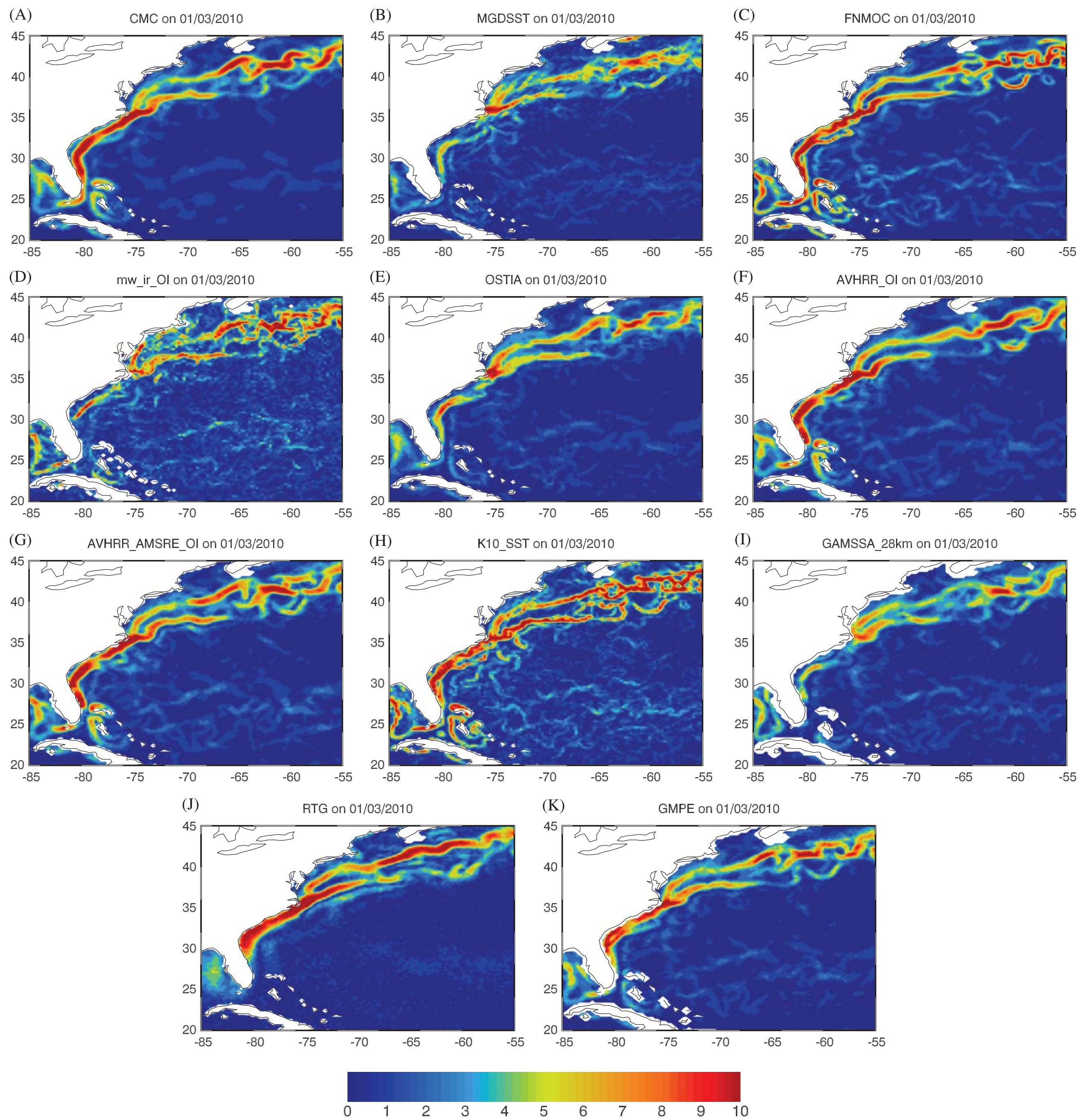

Fig. 3. Averaged east–west, north–south gradients of SST (K 100 km^{-1}) in the Gulf Stream region on 1st March 2010 for (A) CMC, (B) MGDSST, (C) FNMOC, (D) RSS MW/IR, (E) OSTIA, (F) OISST.v2 AVHRR, (G) OISST.V2 AMSR/AVHRR, (H) NAVO K10, (I) GAMSSA, (J) RTG, and (K) GMPE median.

the same day as Fig. 3. This shows that the origin of the median value is not as spatially coherent as might be expected given the horizontal coherence of the SST gradients in GMPE shown in Fig. 3(K).

In order to assess which of the input analyses contributes most to the median, statistics have been calculated over 2010 of the percentage of grid points for which each analysis is the median value. The results are presented in Table 4. This shows that the CMC and OSTIA analyses generally contribute the most globally, with the GAMSSA and NAVO K10 analyses also contributing significantly to the median. It also demonstrates that the analysis contributing to the median is distributed rather than originating from just a few of the analyses all the time. The analysis which is the dominant source of the median varies with latitude. For example the CMC analysis contributes to the median more in the extra-tropics, while OSTIA contributes to the median more in the tropical regions.

4.3. Uncertainty assessments

It would be useful if the GMPE standard deviation could be used as an estimate of uncertainty in the GMPE median since

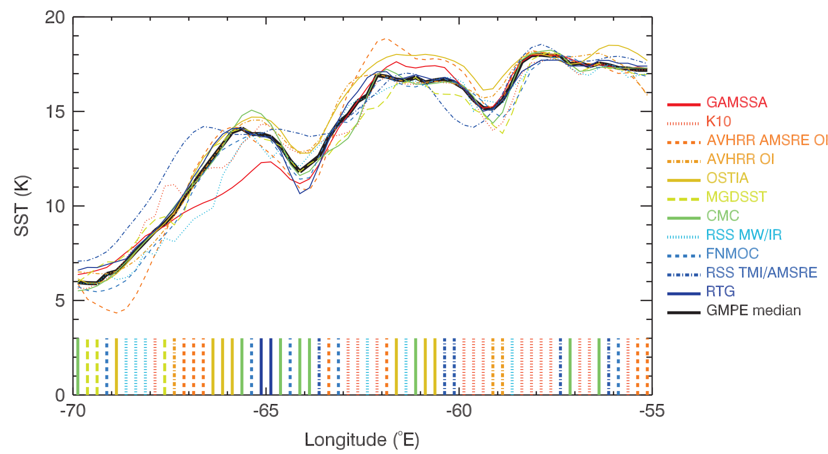

Fig. 4. SST along 40°N between 70°W and 55°W for each of the input analyses and the GMPE median. Shown along the bottom is the origin of the median value (see legend to the right).

Table 4
Percentage of grid points for which each analysis contributes to the median value in GMPE. This is calculated over the whole of 2010 globally and for various latitude bands.

Name of system	*Percentage of points contributing to the median over 2010*			
	Global	*90°S–30°S*	*30°S–30°N*	*30°N–90°N*
CMC	12.9	13.6	12.4	13.1
FNMOC	8.8	9.3	8.7	8.4
GAMSSA	10.3	9.0	11.2	10.2
MGDSST	8.5	8.9	8.3	8.2
NAVO K10	10.1	10.3	9.8	10.4
OISST.v2:AVHRR,	8.2	8.8	7.6	8.5
OISST.v2:AVHRR+AMSR	7.4	7.0	7.3	8.1
OSTIA	12.3	11.6	13.0	11.9
RSS MW	7.2	7.7	7.2	6.2
RSS MW/IR	7.3	7.5	7.0	7.7
RTG	7.1	6.3	7.5	7.3

other, independent estimates of error (such as the one described in Section 4.1) are difficult to obtain in near-real time, and are not easily available before the Argo period.

The spread in the ensemble is due to factors such as different input SST data, and different algorithmic and parameter choices (as described in Section 2). However, one might expect the spread in the ensemble to under-estimate the actual SST error in the median since the analyses are essentially using similar data, and are not radically different from one another in a structural sense. In fact it is not obvious that the spread in the ensemble could be used as a proxy for error since areas of data voids are likely to be common between all of the L4 analyses, resulting in a small spread but larger error.

In order to assess whether the spread can be used as a proxy for error, the differences between the Argo data and the ensemble median (interpolated to the Argo positions) have been averaged for each month during 2010, and over each 10° latitude range for which Argo data are available. The standard deviation of GMPE has been interpolated to the same Argo data locations and these values have then been used to calculate overall standard deviations using a weighted average over the same 10° regions for each month. This procedure therefore provides approximately 150 different comparisons between an error estimate based on independent data, and an error estimate from the ensemble spread. A scatter plot of these values is shown in Fig. 5, together with a line showing a linear fit to the data. The correlation of the two

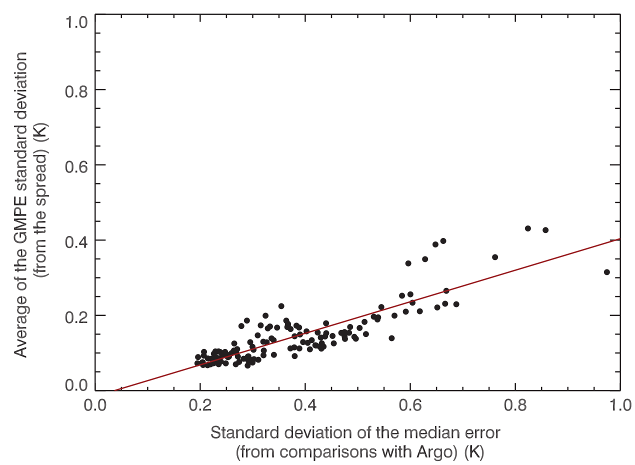

Fig. 5. Scatter-plot comparing two estimates of the error in the GMPE median (K). Along the *x*-axis is the error calculated as the standard deviation of the differences between the GMPE median and near-surface Argo data, along the *y*-axis is the error calculated as the ensemble standard deviation interpolated to the location of the Argo data. Each point represents these values calculated from data over a one-month period in a particular 10° latitude range. The correlation coefficient of these two estimates is 0.87. The linear fit has an intercept at −0.01 and a slope of 0.36.

estimates of the error is 0.87 implying that the GMPE standard deviation does have a strong relationship with the actual error in the median value. The line of linear fit has a gradient of 0.36 indicating that the ensemble standard deviation under-estimates the actual error by more than 60%, and this value could be used for all ranges of the error to provide an estimate of the actual error standard deviation. However, this is only likely to be valid over the large space and time scales assessed here—locally, the spread in the ensemble may not represent the error.

5. Summary

A description of the L4 systems contributing to the GHRSST has been provided with an overview of the various stages in the production of the L4 analyses, and some of the differences between the analyses highlighted. The GMPE system has also been described and its usefulness in terms of illustrating differences between the L4 analyses, and therefore acting to facilitate their development, has been shown.

The ability of the GMPE median to provide an accurate and unbiased estimate of global foundation SST has been assessed using near-surface Argo data. It has been shown that the median value provides a more accurate estimate of SST than any of the contributing L4 analyses with a standard deviation error of 0.40 K and a global bias of 0.03 K. The horizontal gradients in the GMPE median have been compared with the various input analyses and shown to be spatially coherent, with gradients within the range of those from the contributing analyses.

The possibility of using the ensemble standard deviation as a proxy for error in the median has also been explored. This showed that, at least over large scales, the spread in the ensemble is well-correlated with the analysis error, and that the standard deviation under-estimates the analysis error by a factor of about 0.36 for all regions and times investigated. An assessment of this relationship at smaller spatial scales would be useful, although large amounts of independent SST data would be required to carry out such an assessment.

This study has demonstrated that the combination of multiple SST analyses into a multi-product ensemble is a useful activity which can provide both a more accurate estimate of SST and a useful estimate of error. Work is currently underway to develop a version of GMPE for the period 1991 to present, including all the available analyses during that period. An assessment of the number of analyses required to produce a useful ensemble will be required to do this (since only four of the GHRSST L4 systems currently go back to 1991). It would also be useful to assess whether weighting each input analysis based on its relative accuracy, using the comparisons with Argo data described in this paper, could provide a more accurate analysis than the simple median calculation described here.

Acknowledgments

The authors would like to thank John Stark who initially designed and implemented the GMPE system. More recently, the research leading to these results has received funding from the European Community's Seventh Framework Program FP7/2007-2013 under Grant agreement no. 218812 (MyOcean). A. Ignatov and P. Dash acknowledge the support from the Joint Center for Satellite Data Assimilation (JCSDA) Science and Development Implementation (JSDI) program.

References

Beggs, H., Zhong, A., Warren, G., Alves, O., Brassington, G., Pugh, T., 2011. RAMSSA—an operational, high-resolution, multi-sensor sea surface temperature analysis over the Australian region. Aust. Meteorol. Oceanogr. J. 61, 1–22.

Brasnett, B., 1997. A global analysis of sea surface temperature for numerical weather prediction. J. Atmos. Oceanic Technol. 14, 925–937.

Brasnett, B., 2008. The impact of satellite retrievals in a global sea-surface-temperature analysis. Q. J. R. Meteorol. Soc. 134, 1745–1760.

Cummings, J.A., 2005. Operational multivariate ocean data assimilation. Q. J. R. Meteorol. Soc. 131, 3583–3604.

Cummings, J.A., 2011. Ocean data quality control. In: Schiller, A., Brassington, G. (Eds.), Operational Oceanography in the 21st Century. Springer, pp. 91–122.

Dash, P., Ignatov, A., Kihai, Y., Sapper, J., 2010. The SST quality monitor (SQUAM). J. Atmos. Oceanic Technol. 27, 1899–1917.

Dash, P., et al. Group for High Resolution SST Analysis Fields Inter-comparisons. Part 2: The Level 4 SST Quality Monitor (L4-SQUAM). Deep-Sea Res. II, this issue [http://dx.doi.org/10.1016/j.dsr2.2012.04.002].

Donlon, C., et al., 2007. The global ocean data assimilation experiment high-resolution sea surface temperature pilot project. Bull. Am. Meteorol. Soc. 88, 1197–1213. (http://dx.doi.org/10.1175/BAMS-88-8-1197).

Donlon, C., Casey, K., Robinson, I., Gentemann, C., Reynolds, R., Barton, I., Arino, O., Stark, J., Rayner, N., Le Borgne, P., Poulter, D., Vazquez-Cuervo, J., Armstrong, E., Beggs, H., Llewellyn-Jones, D., Minnett, P., Merchant, C., Evans, R., 2009. The GODAE high-resolution sea surface temperature pilot project. Oceanography 22, 34–45.

Donlon, C.J., Martin, M., Stark, J.D., Roberts-Jones, J., Fiedler, E., Wimmer, W., 2012. The operational sea surface temperature and sea ice analysis (OSTIA) system. Remote Sensing Environ. 116, 140–158, http://dx.doi.org/10.1016/j.rse.2010.10.017.

Gemmill, W., Katz, B., Li, X., 2007. Daily Real-Time, Global Sea Surface Temperature High-Resolution Analysis: RTG_SST_HR. Technical Note No. 260, p. 2007.

Gentemann, C.L, Donlon, C.J., Stuart-Menteth, A., Wentz, F.J., 2003. Diurnal signals in satellite sea surface temperature measurements. Geophys. Res. Lett. 30, 1140–1143, http://dx.doi.org/10.1029/2002GL016291.

Gentemann, C.L., Wentz, F.J., DeMaria, M., 2006. Near Real Time Global Optimum Interpolated Microwave SSTs: Applications to Hurricane Intensity Forecasting. Presented at 27th Conference on Hurricanes and Tropical Meteorology, Monterey, CA.

GHRSST Science Team, 2010. The Recommended GHRSST Data Specification (GDS) Revision 2.0 Technical Specifications. The GHRSST International Project Office. Available from: 〈http://www.ghrsst.org〉.

Ingleby, B., Huddleston, M., 2007. Quality control of ocean temperature and salinity profiles—historical and real-time data. J. Mar. Syst. 65, 158–175.

Kurihara, Y., Sakurai, T., Kuragano, T., 2006. Global daily sea surface temperature analysis using data from satellite microwave radiometer, satellite infrared radiometer and in-situ observations. Weather Bull. 73, 1–18. (in Japanese).

Le Borgne, P., Roquet, H., Merchant, C.J., 2011. Estimation of SST from the spinning enhanced visible and infrared imager, improved using numerical weather prediction. Remote Sensing Environ. 115, 55–65.

Le Traon, P.-Y., 2011. Satellites and operational oceanography. In: Schiller, A., Brassington, G. (Eds.), Operational Oceanography in the 21st Century. Springer, pp. 29–154.

Liang, X., Ignatov, A., Kihai, Y., 2009. Implementation of the community radiative transfer model (CRTM) in AVHRR clear-sky processor for oceans and validation against nighttime radiances. J. Geophys. Res. 114, D06112, http://dx.doi.org/10.1029/2008JD010960.

Maloney, E.D., Chelton, D.B., 2006. An assessment of the sea surface temperature influence on surface wind stress in numerical weather prediction and climate models. J. Climate 19, 2743–2762.

Martin, M.J., 2011. Ocean forecasting systems: product evaluation and skill. In: Schiller, A., Brassington, G. (Eds.), Operational Oceanography in the 21st Century. Springer, pp. 611–631.

Maturi, E., Harris, A., Mittaz, J., Merchant, C., Potash, B., Meng, W., Sapper, J., 2008. NOAA's sea surface temperature products from operational geostationary satellites. Bull. Am. Meteorol. Soc. 89, 1877–1888.

Merchant, C.J., Le Borgne, P., Roquet, H., Marsouin, A., 2008. Optimal estimation of SST from split-window observations. Remote Sensing Environ. 112, 2469–2484.

Merchant, C.J., Le Borgne, P., Roquet, H., Marsouin, A., 2009. SST from a geostationary satellite by optimal estimation. Remote Sensing Environ. 113, 445–457.

Petrenko, B., Ignatov, A., Shabanov, N., Kihai, Y., 2011. Development and evaluation of SST algorithms for GOES-R ABI using MSG SEVIRI as a proxy. Remote Sensing Environ. 115, 3647–3658, http://dx.doi.org/10.1016/j.rse.2011.09.003.

Rayner, N.A., Parker, D.E., Horton, E.B., Folland, C.K., Alexander, L.V., Rowell, D.P., Kent, E.C., Kaplan, A., 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J. Geophys. Res. 108 (D14), 4407, http://dx.doi.org/10.1029/2002JD002670.

Reynolds, R.W., Smith, T.M., Liu, C., Chelton, D.B., Casey, K.S., Schlax, M.G., 2007. Daily high-resolution blended analyses for sea surface temperature. J. Climate 20, 5473–5496.

Reynolds, R.W., Chelton, D.B., 2010. Comparisons of daily sea surface temperature analyses for 2007–08. J. Climate 23, 3545–3562.

Stark, J.D., Donlon, C., O'Carroll, A., Corlett, G., 2008. Determination of AATSR biases using the OSTIA SST analysis system and a matchup database. J. Atmos. Oceanic Technol. 25, 1208–1217.

Ullman, D.S., Cornillon, P.C., 2000. Evaluation of front detection methods for satellite-derived SST data using in situ observations. J. Atmos. Oceanic Technol. 17, 1667–1675.

Zhong, A., Beggs, H., 2008. Operational Implementation of Global Australian Multi-Sensor Sea Surface Temperature Analysis. Analysis and Prediction Operations Bulletin No. 77. Bureau of Meteorology, Australia, 2 October 2008. 〈http://cawcr.gov.au/projects/SST/GAMSSA_BoM_Operational_Bulletin_77.pdf〉.