Horizon 2020

H2020-EO-2014 NEW IDEAS FOR EARTH-RELEVANT SPACE APPLICATIONS

EUSTACE

(GRANT AGREEMENT 640171)

EU SURFACE TEMPERATURE FOR ALL CORNERS OF EARTH

DELIVERABLE D3.3

Validation report for the intermediate fields inferred from retrievals

Deliverable Title	Validation report for the intermediate fields inferred from retrievals	
Brief Description	*This document reports the results of validation of the EUSTACE intermediate dataset against in situ air temperatures.*	
WP number	3	
Lead Beneficiary	*Darren Ghent, University of Leicester*	
Contributors	*Karen Veal* *Gary Corlett*	
Creation Date Version Number Version Date	03/11/2016 2	
Deliverable Due Date Actual Delivery Date	M20 M23	
Nature of the Deliverable	R	*R – Report*
		DEM – Demonstrator, Pilot, Prototype
		DEC – Dissemination, Exploitation or Communication
		O – Other
Dissemination Level/ Audience	PU	*PU – Public*
		CO - Confidential, only for members of the consortium, including the Commission services

Version	**Date**	**Modified by**	**Comments**
1	17/10/2016	K L Veal/G K Corlett	First draft
2	03/11/2016	KLVeal/G K Corlett	Addressed review comments
3	21/11/2016	KLVeal	Addressed review comments

EUSTACE

Table of Contents

Executive Summary

This document reports the findings of the validation of test versions of EUSTACE intermediate datasets, which comprise estimates of air temperature derived from satellite retrievals of skin temperature, and their uncertainties over land, ocean and ice. Daily mean air temperatures have been estimated over all surfaces, with daily maxima and minima additionally estimated over land. Here we evaluate both the estimated air temperatures and their estimated uncertainties by calculating discrepancies between the satellite-based estimates and independent in situ measurements of air temperature matched in space and time. This test data set covers different periods in each case and our approach, although performed using a common framework, differs slightly for each surface.

The EUSTACE estimates of daily minimum (Tmin) and maximum (Tmax) air temperatures over land are compared with in situ observations from automatic weather stations. The EUSTACE estimates of Tmax were found to agree closely with the in situ data. The median EUSTACE minus in situ discrepancy for the period 2004 to 2012 is -0.06 K with a robust standard deviation (RSD) of 3.40 K. The EUSTACE Tmin over land has a small negative bias with respect to the in situ data. The median discrepancy for Tmin is -0.54 K, with a robust standard deviation (RSD) of 3.48 K. The EUSTACE uncertainty estimates were consistent in the majority of cases (to within a few tenths of a Kelvin) with discrepancies seen between air temperature estimates and reference measurements; taking into account both uncertainties in the reference measurements themselves and uncertainties arising from the process of matching point to pixel data.

EUSTACE estimates of daily mean (Tmean), minimum, and maximum air temperature over land ice and sea ice are compared with coincident observations from land-based weather stations and sea ice buoys in both the northern and southern hemispheres. EUSTACE estimates were found to have small, mostly warm, biases when compared to in situ data. Median discrepancies for Tmean were less than 1 K in magnitude, ranging from -0.88 to 0.80 K with RSDs of 2.52 – 3.43 K depending on hemisphere and land/sea/marginal ice zone. Again dependent on hemisphere and domain, EUSTACE Tmax estimates had median discrepancies of -1.20 to 1.28 K (RSDs between 1.49 and 3.49 K) and Tmin estimates had median discrepancies of -1.39 to 1.24 K (RSDs between 3.38 and 5.23). The EUSTACE uncertainty estimates were found to be lower than model estimates in most cases. Better agreement was found for estimates of uncertainty in Tmean and Tmax in the northern

hemisphere marginal ice zone and for uncertainty in Tmax over northern hemisphere land ice.

EUSTACE estimates of daily mean air temperature are compared with night-time marine air temperature measurements from ships. EUSTACE estimates of marine air temperature in this test dataset are biased warm by 0.3-0.4 K (as compared to a smaller warm bias of 0.0-0.2 K in the previous version). Cold biases are observed over the major western boundary currents. Warm biases are seen at the northern and southern edges of the data. The EUSTACE marine air temperature uncertainty discrimination and magnitude is excellent up to values of around 4 K and is dominated by the locally correlated uncertainty term.

1. Validation of EUSTACE land surface air temperature test dataset

1.1. Data

1.1.1 EUSTACE Land Intermediate Dataset

This analysis focuses on the second version of the EUSTACE intermediate air temperature dataset over land (referred to as Test Dataset 2, TD2). TD2 is also compared with a previous version of the dataset (produced earlier in the project) which is here referred to as Test Dataset 1 (TD1).

The EUSTACE test datasets provide estimates of daily minimum (Tmin) and daily maximum (Tmax) surface air temperature over land in daily files for the period starting 04/07/2002 and ending 31/12/2012. Air temperature is estimated from the MODIS Aqua skin temperature retrievals via a regression-based method. Data are provided on equal angle longitude-latitude grids with a spatial resolution of 0.25°. Grid box centre coordinates run from 89.875°S to 89.875°N in latitude and from 179.875°W to 179.875°E in longitude.

Uncertainties on the temperature estimates are provided at the grid box level. Three components are given: random, correlated and systematic. It has been assumed in this work that the total uncertainty on a EUSTACE grid box estimate can be calculated by adding these three terms in quadrature.

1.1.2 In situ reference data

In situ weather station surface air temperature data are used to validate the EUSTACE air temperatures. Data from the stations in the Atmosphere Radiative Measurement (ARM) network and US Climate Reference Network (USCRN) were obtained from the EUSTACE Match-up Database (EU_MDB). The EU_MDB contains station measurements along with land cover information, from modified Globcover data [Ghent, 2012], in daily files for each station. A total of 233 stations were available: 7 from the ARM network and 226 from USCRN. These in-situ stations covered a wide variety of terrain, elevations and biomes but are dominated by locations in the contiguous U.S.A. Stations used to validate the EUSTACE air temperature estimates are independent from those used to develop the skin/air temperature regression.

1.2. Approach

The EU_MDB was used to select all in situ 2 m air temperatures that were coincident with the EUSTACE TD2 data: i.e. those occurring on the same day (00:00 to 24:00 local solar time) and in the same 0.25° grid box as the matching TD2 data point. The in situ minimum and maximum temperatures were found for each matchup station-day.

Thus, two pairs of data were available for comparison: the EUSTACE TD2 Tmin and in situ Tmin; and the EUSTACE TD2 Tmax and in situ Tmax. TD2 data are unavailable at some stations, on some days (most likely due to cloud cover) so no comparison could be performed for these data points.

Note, in situ measurements are effectively point measurements and the EUSTACE air temperatures are estimates of the mean value over a 0.25° x 0.25° grid box which is equivalent to approximately 25 km x 25 km at the equator. We are, therefore, assuming that the in situ measurement is truly representative of the EUSTACE grid box. This assumption introduces an uncertainty into the matchup process which is likely to be larger in grid boxes where land cover, elevation and topography are more variable. Validation of the final EUSTACE dataset will make use of observations from the Global Historical Climate Network (Daily). The number of stations used will be greater in number and cover a wider geographical area. It is likely that there will be more than one station in some EUSTACE gridboxes which may necessitate a strategy different to the current one station to one gridbox match-up process.

The matched temperature data are stored for further analysis, along with coordinate data, land cover type and elevation at the in situ station. The analysis considers discrepancies between the matched data calculated as EUSTACE test dataset minus in situ temperature differences. Robust statistics are used with normal distributions characterized by the median and a robust standard deviation (RSD) which is calculated as 1.4826 times the median absolute deviation from the median.

1.3. Results

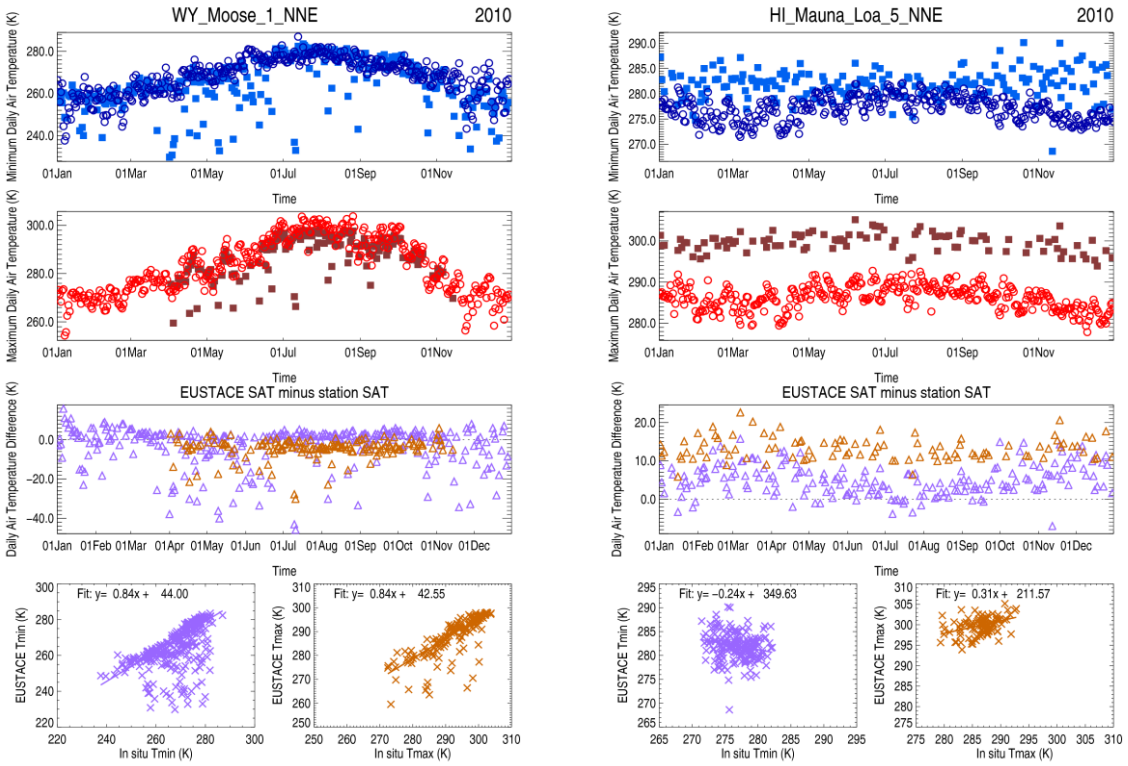
1.3.1 Daily Tmax and Tmin

Figure 1-1 shows example time series plots for 2010 at Moose, Wyoming and Mauna Loa, Hawaii. The availability of EUSTACE data at Mauna Loa is approximately constant throughout the year. In contrast, at Moose, the EUSTACE dataset has no Tmax data during

the winter (December to April). This may be a result of daytime cloud cover or overzealous cloud clearing of the satellite data.

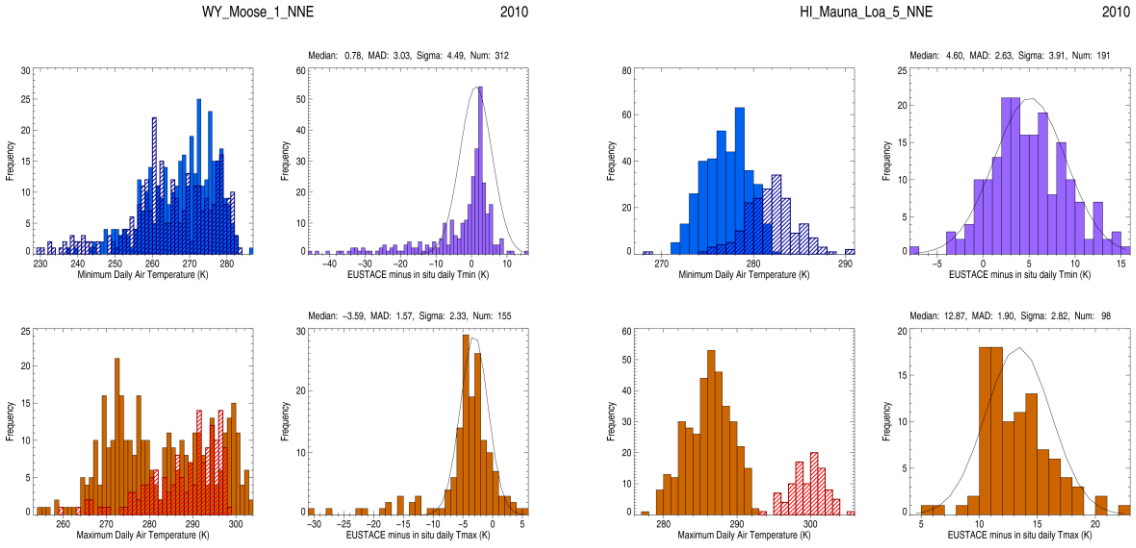
The EUSTACE data represents the seasonal cycle well at Moose, WY. However, there are several EUSTACE data that exhibit large (order of magnitude 10 K) cold discrepancies when compared to the in situ data. These large cold discrepancies are more prevalent in the Tmin data but also occur to a lesser extent in the Tmax time series. A possible explanation is that the extreme cold biases are caused by cloud contamination in the satellite inputs to the EUSTACE data. These data are seen as an extended 'cold' tail in histograms of the match-up discrepancies (Figure 1-2) and the discrepancy RSD is consequently large. The median discrepancy in Tmin is +0.78 K with an RSD of 4.49 K. In addition to the problem of cloud contamination, the data at Moose demonstrates a cold bias in the Tmax with a median discrepancy of -3.59 K (and RSD 2.33 K).

The large cold discrepancies are largely absent from the Tmin and Tmax time series at Mauna Loa. The EUSTACE data for Mauna Loa exhibits warm discrepancies which are fairly systematic for Tmax. The Mauna Loa station is situated at 3397 m above sea-level, in a region of highly variable topography. The discrepancies may be due to the station not being truly representative of the EUSTACE gridbox and/or due to biases in the EUSTACE data associated with the high altitude. The Tmin discrepancies are smaller and are largest during winter months when temperatures are cooler. At Mauna Loa, the median bias is +4.60 (+12.87) K for Tmin (Tmax) with an RSD of 3.91 (2.82) K.



(a) (b)

Figure 1-1: Match-up data for 2010 at (a) Moose, Wyoming and (b) Mauna Loa, Hawaii. EUSTACE test data are represented by filled squares and the in situ data are represented by open circles. Minimum temperatures are given in blue (top, each panel) and maximum temperatures are coloured red/brown (second from top). EUSTACE test data minus in situ discrepancy is plotted (second from bottom) for maximum daily temperature (mauve) and minimum daily temperature (tan). The bottom row in each panel shows scatter plots of EUSTACE test data against in situ temperatures.

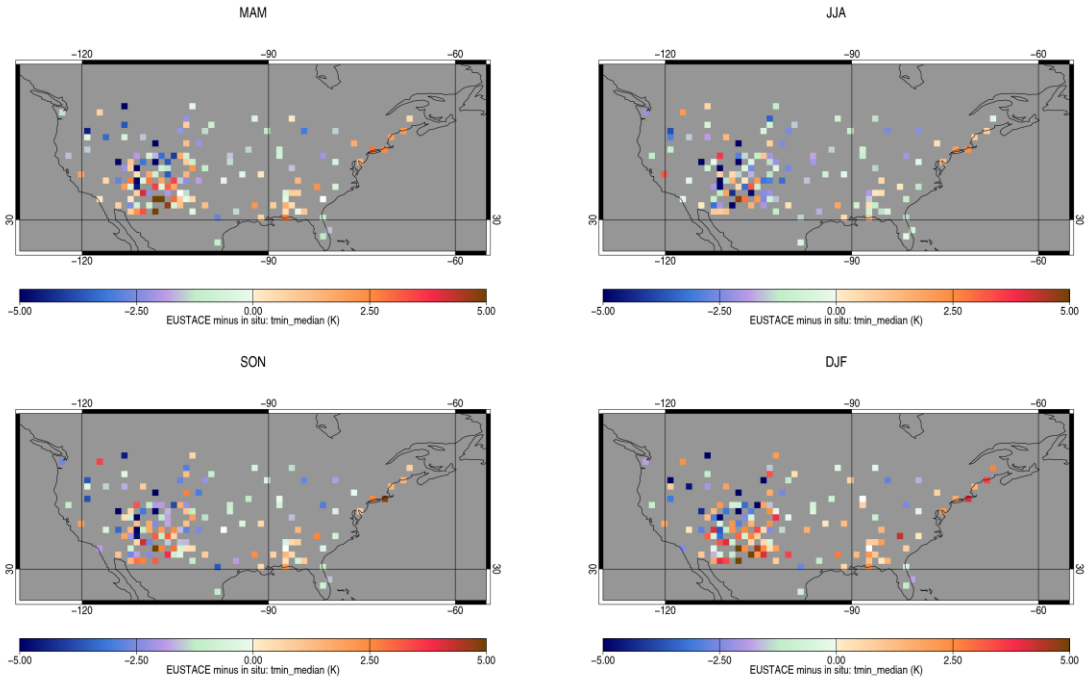


(a) (b)

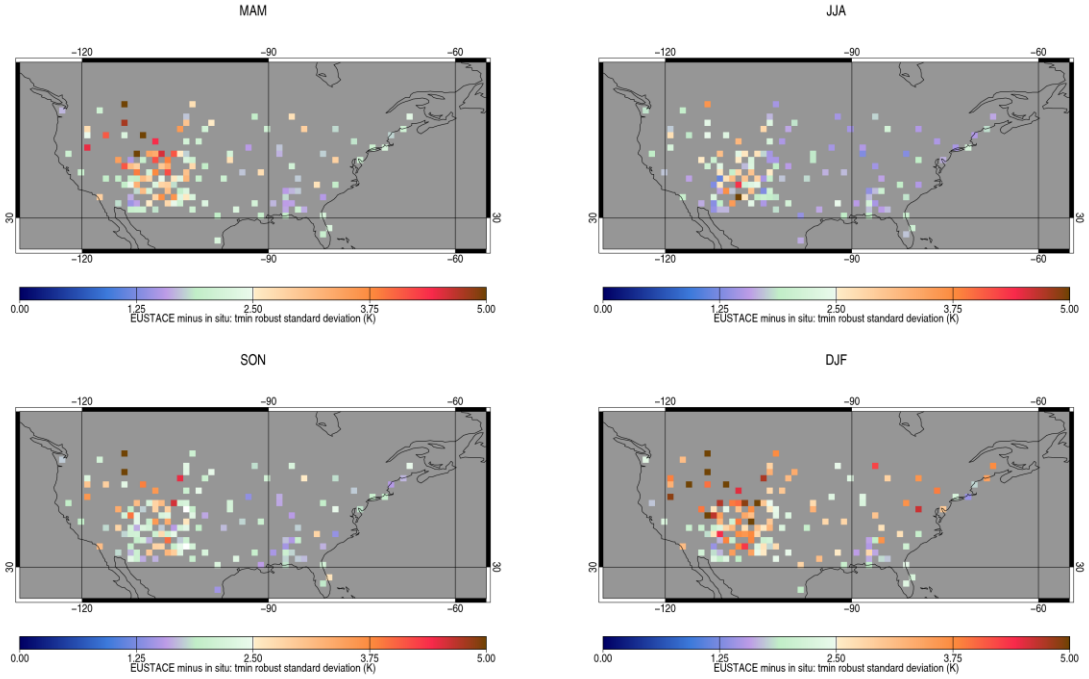
Figure 1-2: Histograms of EUSTACE test data and in situ daily maximum and minimum temperatures for 2010 at (a) Moose, Wyoming and (b) Mauna Loa, Hawaii. In situ distributions are plotted in solid colours (Tmin in blue, Tmax in tan) and hatched bars show EUSTACE distributions (Tmin in blue, Tmax in red). Right-hand plots of each block of four are histograms of match-up discrepancy values (Tmin in mauve, Tmax in tan); a gaussian function calculated using the median, RSD and modal value of the distribution (as the centre, standard deviation

and maximum value, respectively, of the gaussian function) is over plotted. Histogram bin size, in all cases, is 1 K.

Figure 1-3 and Figure 1-4 show maps of seasonal median and RSD in matchup discrepancy over the period 2004 to 2012 for Tmin and Tmax respectively. Tmin discrepancies have a northwest to southeast gradient across the USA with negative values in the northwest and positive values in the southeast. The large negative discrepancies tend to occur in mountainous regions with highly variable topography and elevation, sparse or no vegetation and seasonal snow cover. Discrepancies are colder in the summer months (JJA) than in other seasons. For Tmax, the largest seasonal mean discrepancies (both positive and negative) are again seen towards the west of the region, in mountainous terrain. However, large positive discrepancies (> 3 K) are also seen in the winter (DJF) at more northern latitudes and also at stations along the north-eastern seaboard. Seasonal changes are apparent in the discrepancy RSDs for both Tmin and Tmax. RSDs are smallest in summer (JJA) and largest in winter (DJF) and spring (MAM). In general the RSDs increase with increasing latitude and the Tmin RSDs tend to be larger than the Tmax RSDs.

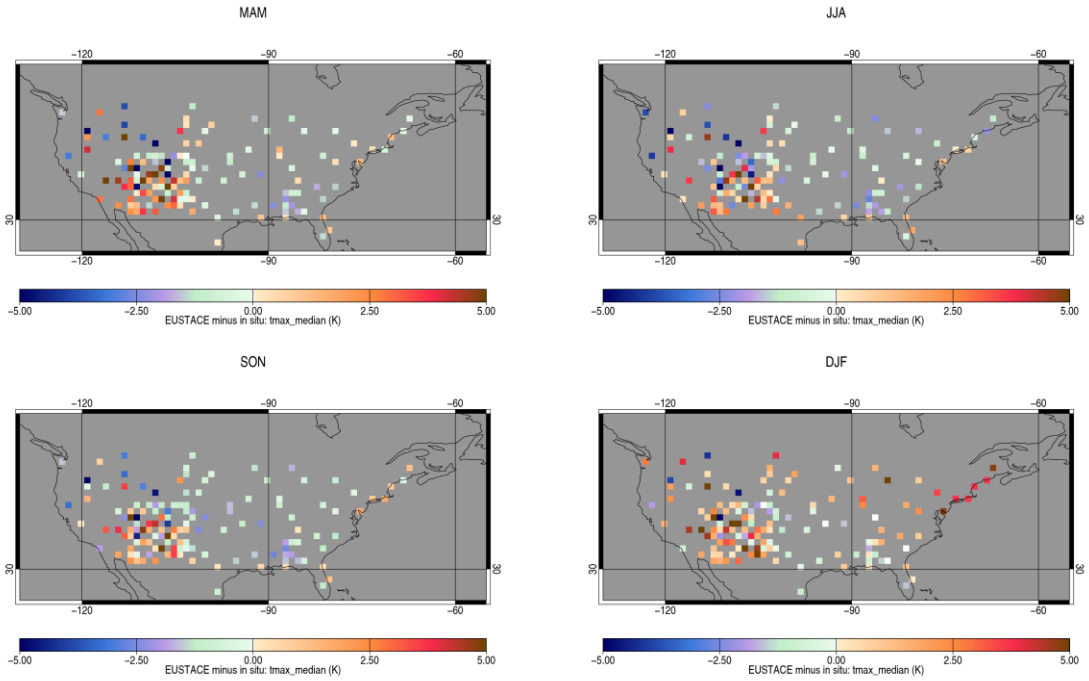


(a)

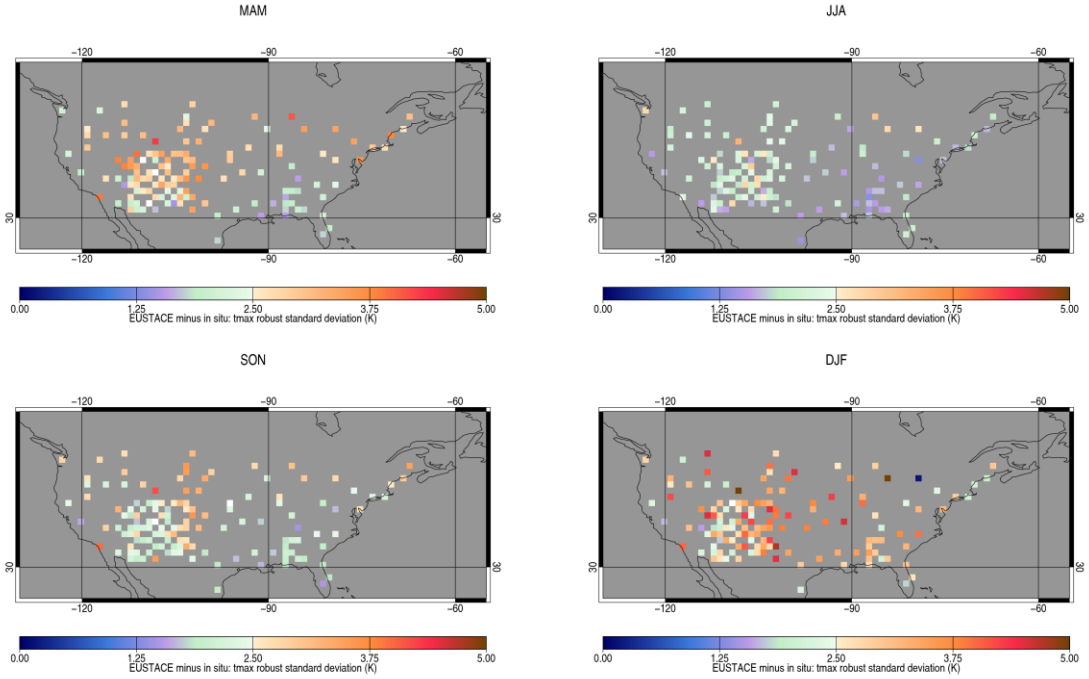


(b)

Figure 1-3: Seasonal (a) median and (b) robust standard in discrepancy between EUSTACE and in situ daily minimum temperature for 2004 – 2012. In each panel: March-April-May (top left), June-July-August (top right), September-October-November (bottom left), December-January-February (bottom right). Data have been averaged onto a 1° x 1° grid to improve clarity.



(a)



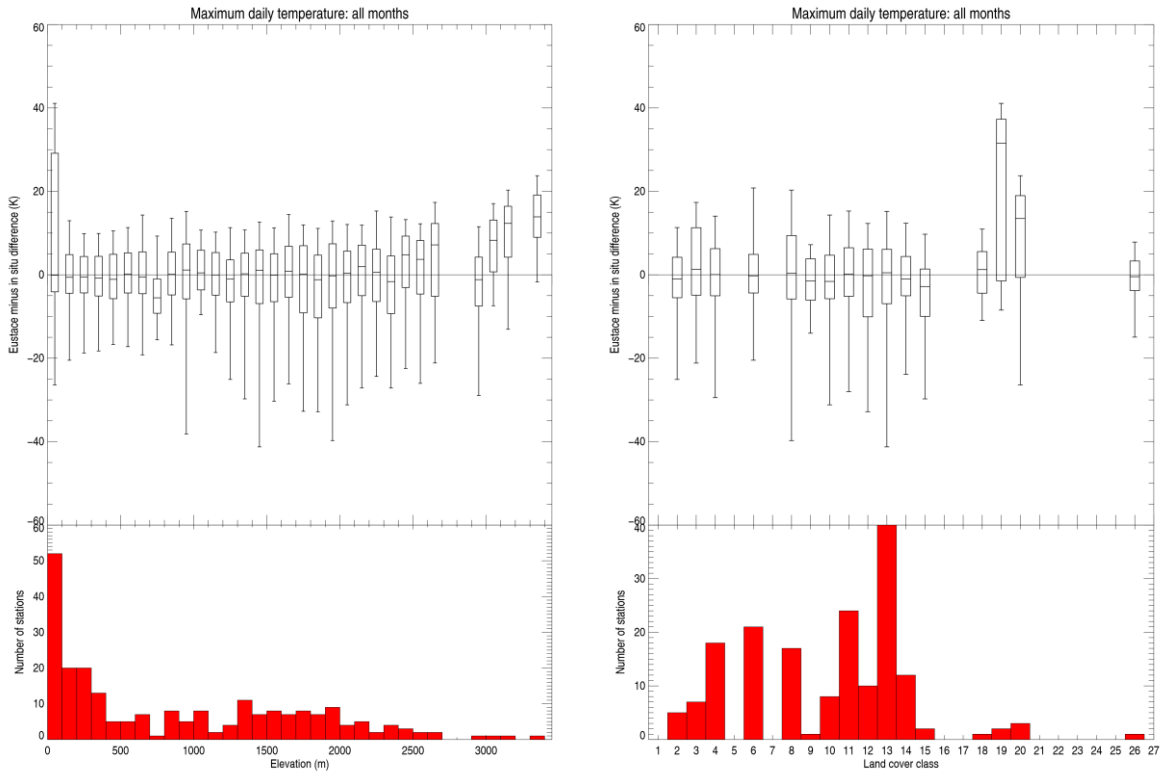
(b)

Figure 1-4: Seasonal (a) median and (b) median absolute discrepancy in daily maximum temperature for 2004 - 2012 (see caption to Figure 1-3 for legend).

At locations with elevations above 3000 m discrepancies in Tmax are generally positive with 90% of values greater than zero for nearly all seasons (Figure 1-5a). The in situ stations that fall into this group are: Mauna Loa, Hawaii; Socorro, New Mexico; and Beaver, Utah. These positive discrepancies may be due to errors in the EUSTACE data but matchup uncertainties must be considered. These high altitude stations tend to be in terrain with highly variable topography. The temperature at the in situ stations are likely not truly representative of the mean temperature over a EUSTACE grid box. Investigation into the differences between grid box mean elevation and in situ station elevation using a high resolution digital elevation map (DEM) and consideration of temperature lapse rates may, in future work, allow estimation of the uncertainty in the point-to-gridbox matching of temperatures caused by the topographic variability. Median discrepancy in Tmin (**Figure 1-6**) is lower at high altitudes than is the case for Tmax although this may be due to other errors such as cloud contamination of the satellite data offsetting any errors related to high elevation.

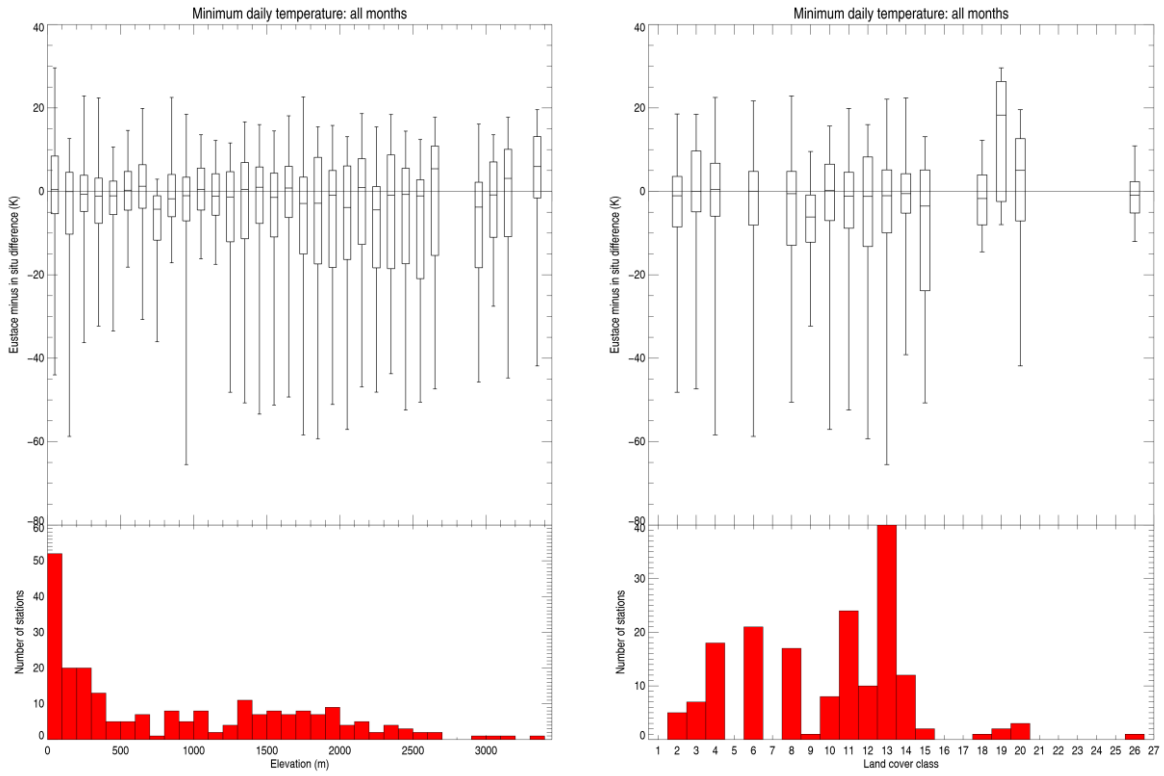
The match-up discrepancy shows little dependence on land cover class (Figure 1-5b). The urban class (biome 19) exhibits a large median discrepancy and spread but this is due to poor quality control of the in situ data from the Darwin station (several aberrant zero values have not been removed). Also, these urban sites may not be representative of the dominant landcover class in the matched EUSTACE gridbox. Biome 20 (bare soils) has a relatively high median discrepancy which is due to the high discrepancies at Mauna Loa discussed above.

Overall, when global values are calculated for the period 2004-2012 (excluding match-ups with Darwin in situ data which includes bad values) then the median discrepancy for Tmin is -0.54 K (RSD 3.48 K) and for Tmax is -0.06 K (RSD 3.40 K). Cloud contamination of the input satellite skin temperature data is the most likely cause of the cold bias and relatively large RSD from reference data of the EUSTACE Tmin. Cloud contamination results in large negative biases in the satellite data which will influence the regression used to model the air temperatures. Cloud clearing of the satellite data is more difficult at night and the modelled Tmin has a greater dependency on the night-time satellite skin temperature than does the modelled Tmax.



(a) (b)

Figure 1-5: Match-up discrepancy statistics for 2004-2012 by in situ site elevation (a) and site land cover class (b) for Tmax. Box and whisker plots show median discrepancy (horizontal bar), fifth and ninety-fifth percentile limits (box) and minimum and maximum values (whiskers). Land cover is taken from a modified version of Globcover data [Ghent, 2012]. Class numbers refer to the following biomes: (1) Post-flood/Irrigated crops, (2) Rainfed crops, (3) Mosaic Cropland/Vegetation, (4) Mosaic Vegetation/Cropland, (5) Broadleaved evergreen and/or semi-deciduous forest, (6) Broadleaved deciduous forest: closed, (7) Broadleaved deciduous forest: open, (8) Needleleaved evergreen forest: closed, (9) Needleleaved deciduous or evergreen forest: open, (10) Mixed broadleaved and needleleaved forest, (11) Mosaic Forest/Shrubland/Grassland, (12) Mosaic Grassland/Forest/Shrubland, (13) Closed to open shrubland, (14) Closed to open grassland, (15) Sparse vegetation, (16) Closed broadleaved forest regularly flooded –Freshwater, (17) Closed broadleaved semi-deciduous and/or evergreen forest regularly flooded –Saline water, (18) Vegetation on regularly flooded or waterlogged soil -Fresh, brackish or saline water, (19) Artificial surfaces (urban areas), (20) Bare soil: other than 21-25, (21) Bare soil: Entisols–Orthents, (22) Bare soil: Shifting sand, (23) Bare soil: Aridisols-Calcids, (24) Bare soil: Aridisols-Cambids, (25) Bare soil: Gelisols-Orthels, (26) Water bodies, (27) Permanent snow and ice.



(a) (b)

Figure 1-6: Match-up discrepancy statistics for 2004-2012 by in situ site elevation (a) and site land cover class (b) for Tmin. See caption to Figure 1-5 for legend.

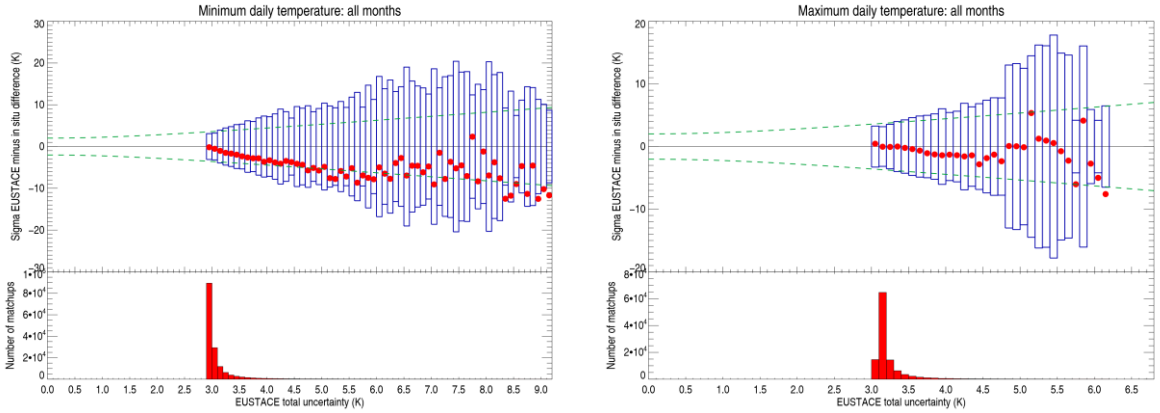
1.3.2 Validation of EUSTACE uncertainty estimates

The method for validation of uncertainties used here is based on that developed for the ESA Climate Change Initiative SST Project [Bulgin et al., 2016]. Variance of matchup discrepancy is modelled as a sum in quadrature of the uncertainties on the in situ measurement ($\sigma_{in\ situ}$), the EUSTACE estimate ($\sigma_{EUSTACE}$), and a term representing the uncertainty inherent in matching a point observation at an in situ station with a 0.25° x 0.25° gridbox mean ($\sigma_{match-up}$):

$$\sigma^2_{discrepancy} = \sigma^2_{in\ situ} + \sigma^2_{match-up} + \sigma^2_{EUSTACE}\ .$$

The 95% confidence limit for automatic weather stations is 0.57 K [ARM, 2011] and this is assumed to be $2\sigma_{in\ situ}$. The match-up uncertainty is assumed to have a global value of 2 K.

Figure 1-7 shows the robust standard deviation of the match-up discrepancy plotted against the EUSTACE estimated total uncertainty (blue bars) along with what the spread in the matchup discrepancy should be if the above model is correct (green dashed line). If the model assumptions are correct and the EUSTACE uncertainties accurate then the model and the match-up discrepancy RSD should agree and the blue bars should lie along the green dashed lines.



(a) (b)

Figure 1-7: Validation of EUSTACE uncertainty, 2004-2012 (a) minimum daily temperature, (b) maximum daily temperature. Green dashed lines show model estimates. Blue bars show match-up robust standard deviations (RSDs). Red dots show median discrepancy in each bin. RSDs are always positive but are also plotted as negative bars to aid comparison with median values.

Figure 1-7a shows that the majority (84%) of the EUSTACE estimates of Tmin in the match-ups have uncertainties between 2.9 and 3.2 K. If we assume the model is correct, then these uncertainties are accurate to within a few tenths of a kelvin. As the EUSTACE uncertainty estimate increases, the disagreement between the model and the match-up discrepancy becomes much larger with the EUSTACE uncertainty generally under-estimated. Also, the median discrepancy becomes more negative with increasing uncertainty, particularly for Tmin. However, note that the sample numbers are small. Figure 1-8 shows the different components of the EUSTACE uncertainty. For the majority of match-ups the size of the EUSTACE uncertainty estimate is dominated by the locally correlated component. The correlated component shows little variation with values in the subset used here, having values confined to between 2.90 and 3.12 K. The random component shows the greatest range and dominates for higher values of the total uncertainty.

The EUSTACE Tmax uncertainties are again in good agreement with the model especially for uncertainties less than 4.8 K (Figure 1-7b). Above 4.8 K, the uncertainties appear to be underestimated most of the time. Again the variation in the uncertainty is dominated by the

random component in most cases. The correlated component is restricted to values between 3.08 and 3.18 K. The random component of the Tmax uncertainty is the component with the greatest range of variation.

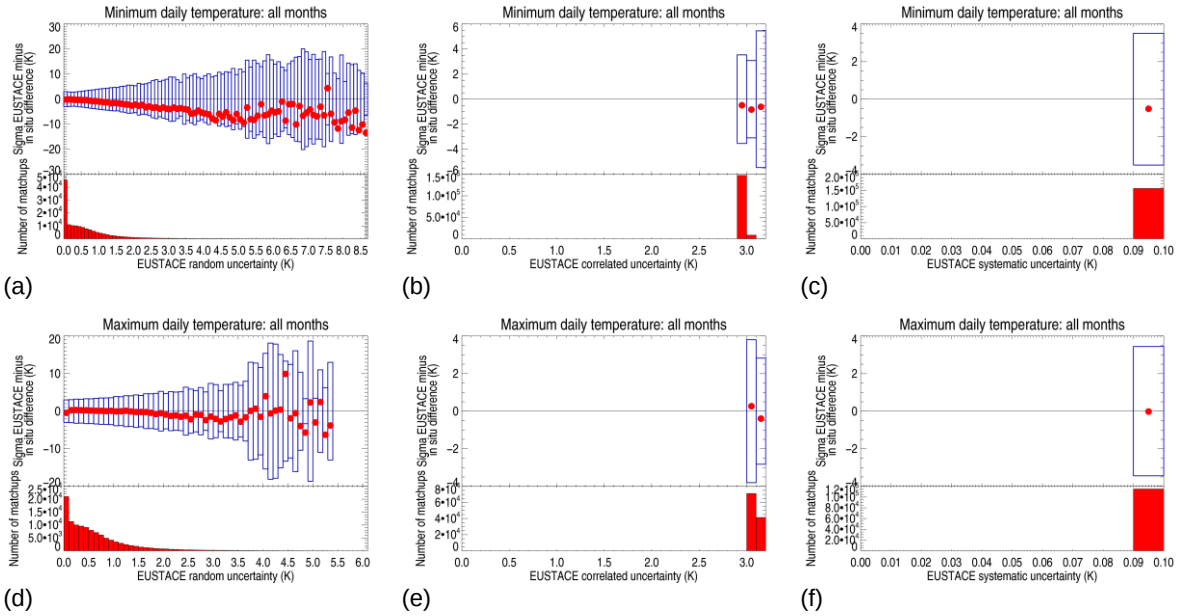


Figure 1-8: Components of the EUSTACE uncertainty for Tmin (a-c) and Tmax (d-f): random component (a, d), locally correlated component (b, e), systematic component (c, f). See caption to Figure 1-7 for explanation of plot components.

1.3.3 Comparison between TD1 and TD2

Test Dataset 1(TD1) consisted of data for 2010 only, so here we compare matchups for TD1 and TD2 for 2010. Global statistics are presented in Table 1-1. The median discrepancy is reduced in magnitude for TD2 compared with TD1; by 20% in the case of Tmin and 85% in the case of Tmax. The robust standard deviations are little changed (< 5%). This indicates that the further development undertaken in EUSTACE since the production of the first test data set has improved the air temperature estimates over land.

Table 1-1: Match-up statistics for 2010 from test dataset 1 (TD1) and test dataset 2 (TD2)

	Tmin			Tmax		
	Median (K)	RSD (K)	No.	Median (K)	RSD (K)	No.
TD1	-0.60	3.41	26932	-0.40	3.26	26932
TD2	-0.48	3.40	30309	-0.06	3.40	22551

The TD1 files contained random, correlated, systematic, and sampling components of uncertainty although the systematic component was zero everywhere. The majority of the TD1 uncertainties lie below 1.5 K and are underestimated by up to a few kelvin (see Figure 1-9). The sampling uncertainty is over estimated in a small number of cases. The TD2 uncertainties show much better agreement with the model estimates.

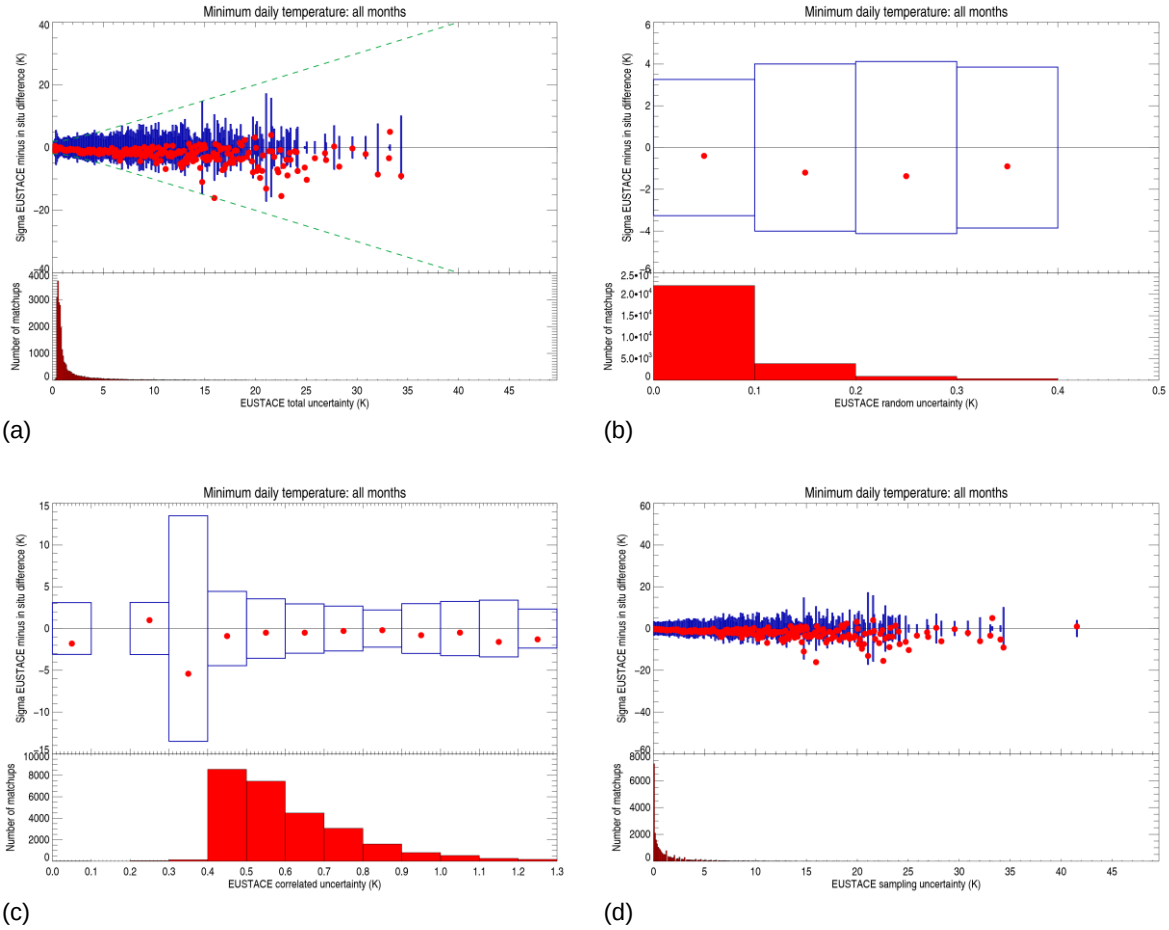


Figure 1-9: Uncertainties for TD1: (a) total uncertainty, (b) random component, (c) locally correlated component, (d) sampling uncertainty. See caption to Figure 1-7 for explanation of plot components.

1.4. Summary

- The global median discrepancy in the second EUSTACE test dataset, for the period 2004-2012, for Tmin is -0.51 K (RSD 3.51 K) and for Tmax is -0.06 K (RSD 3.51 K).

- This analysis uses independent in situ stations largely confined to the contiguous USA. For this region the largest discrepancies between EUSTACE and in situ temperatures are found in western mountainous regions.
- The EUSTACE test data uncertainty estimates show good agreement with the spread of discrepancies between estimated and measured air temperatures, taking into account uncertainties in the reference data and uncertainties arising from matching point to pixel values. For ~84% of the EUSTACE air temperatures evaluated, the estimated uncertainties seem accurate to within a few tenths of a K.
- Version 2 of the test dataset (TD2) shows improvements over version 1 (TD1). TD2 has lower median discrepancy when compared to in situ data than TD1 and the TD2 uncertainties are closer to model estimates.

2.Validation of EUSTACE land ice and sea ice surface air temperature test data set

2.1. Data

2.1.1 EUSTACE Ice Intermediate Dataset (TD2)

Two datasets are provided: northern hemisphere and southern hemisphere air temperatures over ice. Each dataset provides daily estimates of minimum daily temperature, maximum daily temperature, and mean daily temperature. The data provided covered 2007 and 2008. Each temperature estimate is associated with two uncertainty estimates: one which includes a cloud mask uncertainty, and one which does not. The separate components of the uncertainty estimates are also included; random, locally correlated, large-scale systematic, and cloud mask uncertainty components are given with associated correlation length and time scales. The data set also provides masks for land ice, sea ice (sea ice concentration above 30%) and the marginal ice zone (sea ice concentration between 30% and 85%).

Temperature estimates and uncertainties are provided on a 0.25° equal angle latitude longitude grid. Grid points are given at every 0.25° latitude from 60°N (60°S) to 89.75°N (89.75°S) in northern (southern) hemisphere and at every 0.25° longitude between 180.0°W and 179.75°E. Coordinates refer to pixel centres. Note the EUSTACE ice grids differ from the EUSTACE data land grid, being offset by 0.125 ° in both longitude and latitude. Also, the daily estimates are based on a 24 hour period from 00:00 to 24:00 UT unlike the land dataset which uses a local solar day.

This is the first set of test data for the ice-covered regions, so no precursor data set is available for comparison here.

2.1.2 In situ reference data

Lists of in situ stations whose data has been reserved for validation were provided by DMI: data from these stations were not used for training of the ice surface air temperature algorithm. Separate lists were provided for northern hemisphere land ice, northern hemisphere sea ice, southern hemisphere land ice, and southern hemisphere sea ice.

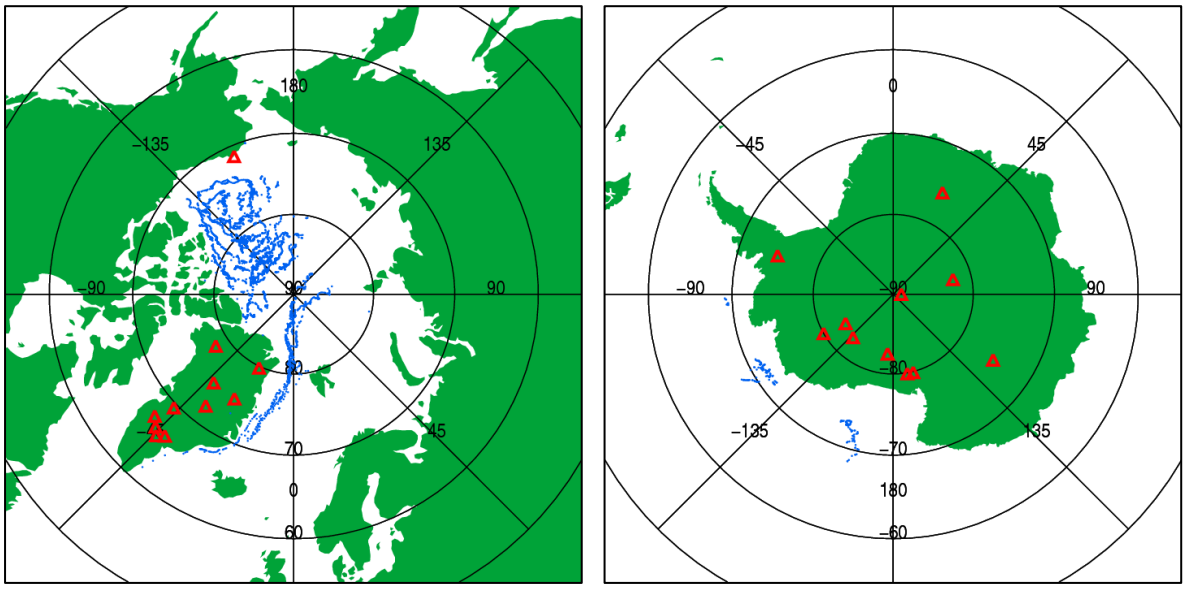
Quality controlled in situ measurements of air temperature were provided by DMI in netCDF files, one file for each station. Air temperatures are provided with time, longitude and latitude

fields and other physical parameters, such as wind speed and relative humidity, amongst others.

2.2. Approach

The approach taken was similar to the method used for creation of the land matchup data with minor differences. In situ daily minimum, maximum and mean daily temperatures at each station were calculated over the UT day to better match the EUSTACE ice test data set estimates (local solar days will be used in future versions). Also, the analysis was done separately for the northern/southern hemispheres and land/sea domains. The EUSTACE test data set contains estimates of minimum, maximum and mean daily temperature along with uncertainty estimates. For brevity, in this report, we concentrate mainly on validation of estimates of the mean daily temperature (Tmean) but show some results for Tmin and Tmax also.

The number of match-ups available for analysis is less than is the case for the land product. This is in part due to the EUSTACE data covering a shorter period (2 years) but also due to the limited availability of in situ data from ice covered regions. Figure 2-1 shows the land-based ice stations with data available during the period and the locations of available in situ sea ice measurements.



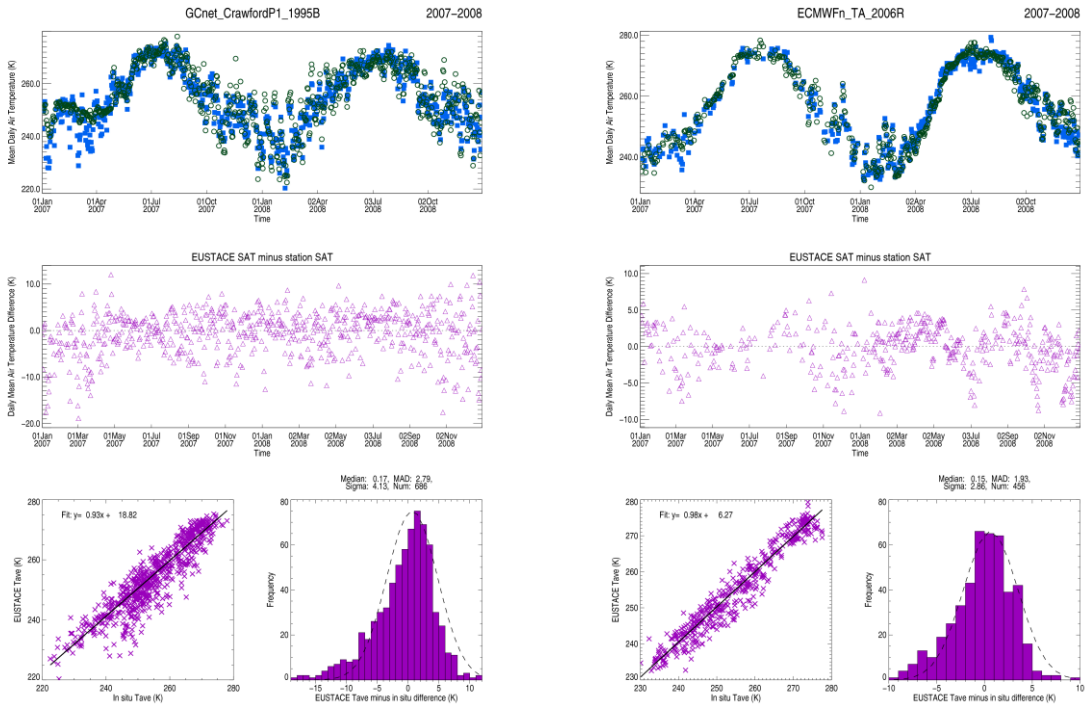
Northern hemisphere Southern hemisphere

Figure 2-1: Locations of in situ data available for validation of EUSTACE ice data for 2007-2008 period. Red triangles show location of land stations, blue dots show locations associated with sea-ice buoy data points that are coincident with EUSTACE mean air temperature estimates.

2.3. Results

2.3.1 Daily mean temperature (Tmean)

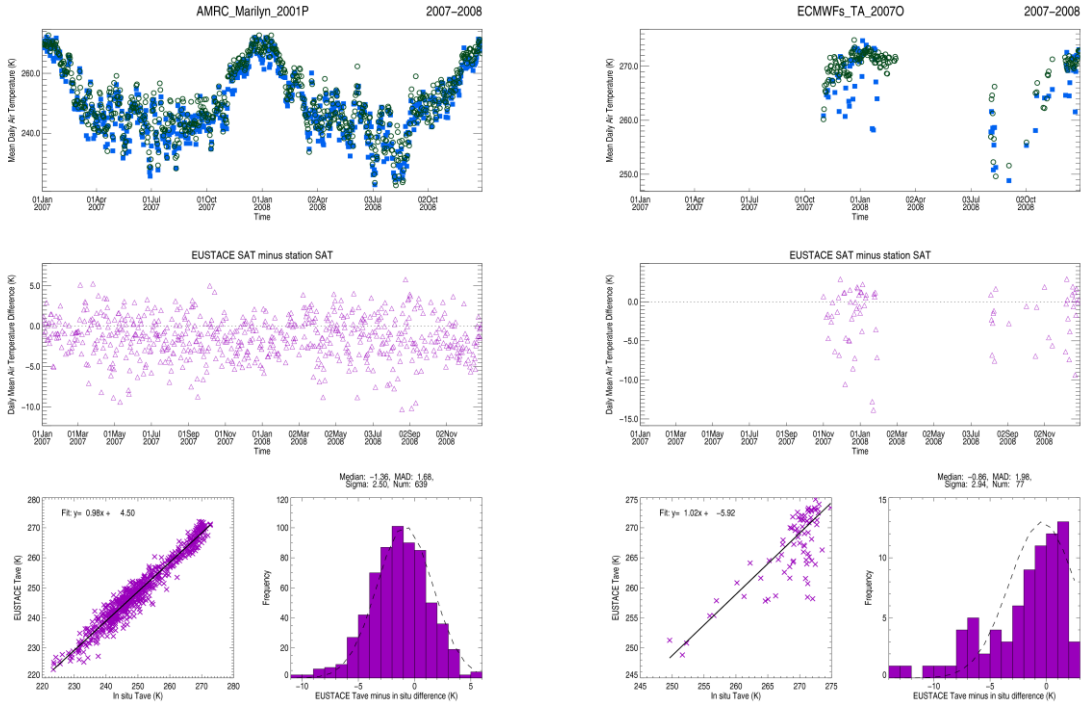
Time series of Tmean from example stations, one from each of the four regions, are shown in Figure 2-3 and Figure 2-3. The EUSTACE data captures both the seasonal and shorter term variability well.



Land ice northern hemisphere Sea ice northern hemisphere

Figure 2-2: Time series of match-up data for 2007-2008 at example in situ locations in the northern hemisphere. In each panel: time series of Tmean (top) where EUSTACE data are represented by blue filled squares and the in situ data are represented by green open circles; time series of EUSTACE minus in situ matchup discrepancy (middle); scatter plot of EUSTACE Tmean against in situ Tmean (bottom left); and (bottom right) a histogram of matchup discrepancy with a gaussian function calculated using the median, RSD and modal value of the distribution (as the centre, standard deviation and maximum value, respectively, of the gaussian function) over-plotted.

Match-up statistics for each region, for 2007-2008 are shown in Table 2-1. The marginal ice zone (between 30% and 85% ice concentration) has been analyzed separately from a sea ice zone with greater than 85% sea ice concentration. Note, very few match-up pairs are available for southern hemisphere sea-ice.



Land ice southern hemisphere

Sea ice southern hemisphere

Figure 2-3: Time series of match-up data for 2007-2008 at example in situ locations in the southern hemisphere. For legend see caption to **Figure 2-2**.

The regional median discrepancies are mostly positive and small (less than 1 K in most cases) meaning that in most regions the EUSTACE data are biased warm relative to the in situ measurements. The largest median discrepancies occur for the northern hemisphere Tmin, in the land ice (1.24 K) and marginal ice zones (-1.39 K) and in the southern hemisphere Tmax, in the land ice (1.28 K) and sea ice zones (-1.20 K). The southern hemisphere sea ice zone has negative median biases for all three variables (Tmean, Tmin and Tmax) although the bias for Tmin is quite small at -0.07 K. The largest RSDs occur for Tmin with values greater than 3 K (up to 5.3 K) in all regions whereas values for Tmax and Tmean are less than 3.5 K in all regions. Cloud contamination of the satellite data is a likely cause as it will cause erroneously low satellite ice temperatures (ISTs) to be selected as the minimum ISTs which are input to the EUSTACE air temperature model.

Table 2-1: Regional match-up statistics for EUSTACE minus in situ discrepancy in daily minimum and daily maximum temperature, for 2007-2008. Matchups are confined to those where the OSI-SAF ice concentration is greater than or equal to 75%. Robust standard deviations (RSD) are calculated by scaling the median absolute deviation from the median.

Region	Discrepancy in Tmean			Discrepancy in Tmin			Discrepancy in Tmax		
	Median (K)	RSD (K)	No.	Median (K)	RSD (K)	No.	Median (K)	RSD (K)	No.
Northern hemisphere land ice	0.74	3.43	4723	1.24	4.23	3919	0.12	2.53	3919
Northern hemisphere sea ice (>85%)	0.61	3.29	4251	0.84	4.37	3205	0.76	3.53	3232
Northern hemisphere MIZ	0.39	2.77	738	-1.39	5.23	364	0.20	2.74	364
Southern hemisphere land ice	0.80	2.71	5252	0.94	3.39	4010	1.28	3.49	3596
Southern hemisphere sea ice (>85%)	-0.88	2.79	118	-0.07	3.58	71	-1.20	2.97	70
Southern hemisphere MIZ	-0.01	2.52	99	-0.22	3.38	42	0.92	1.49	38

2.3.2 EUSTACE uncertainty estimates

The method for validation of the EUSTACE ice uncertainties and interpretation of plots is the same as for EUSTACE land data and is described in Section 1.3. Although the model used is the same as that used for land, the model parameters used for land ice and sea ice are different (see Table 2-2). The in situ measurement uncertainty is taken from the value for Automatic Weather Stations [ARM, 2011] over land (given as a 95% confidence interval and here assumed a 2 sigma uncertainty) but is assumed higher for sea ice buoys. The match-up uncertainty for ice is assumed less than for land because land cover over ice shows less

variability. For grid boxes which are not entirely ice covered the value used may be an underestimate.

Table 2-2: Parameters for discrepancy uncertainty model over ice regions

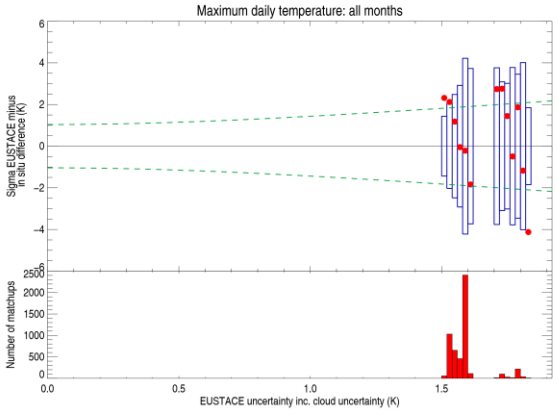
Model parameter	Land ice (K)	Sea ice (K)
In situ uncertainty ($\sigma_{in\ situ}$)	0.57/2. [ARM, 2011]	1.0
Match-up uncertainty ($\sigma_{match\text{-}up}$)	1.0	1.0

Figure 2-4 and Figure 2-5 shows the validation of the EUSTACE ice total uncertainty (including a cloud mask uncertainty component) for Tmean and Tmax, respectively, in the six regions. Within each region, all uncertainty values associated with match-up pairs fall into a narrow range of width a few tenths of a kelvin. The low sample numbers, coupled with the relatively narrow range of uncertainty values, makes interpretation of the uncertainty analysis difficult, more so since in some regions over 75% of values fell within one histogram bin of width 0.02 K. The narrow range of values also restricts our ability to assess the accuracy of the model used to simulate the matchup discrepancy RSD.

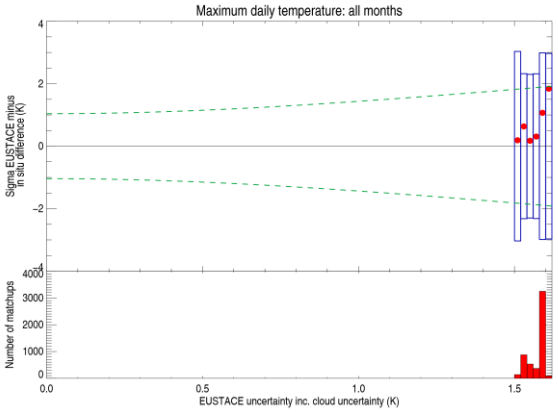
With the above caveats, we attempt some interpretation. The EUSTACE Tmean uncertainty estimates lie within a range 1.5 – 1.9 K. The estimates are generally too low for land ice, northern hemisphere sea ice, and southern hemisphere marginal ice. The sample numbers for southern hemisphere sea ice are too low to enable any conclusions to be drawn. However, the results for the northern hemisphere marginal ice zone show approximately 55% of EUSTACE uncertainty estimates agree with the modelled matchup discrepancy RSD.

For Tmax, the EUSTACE uncertainty estimates lie in a range from 1.98 K to 2.22 K. The EUSTACE uncertainty estimates for Tmax are generally too low over southern hemisphere land ice, northern hemisphere sea ice and the southern hemisphere MIZ. Estimates over northern hemisphere land ice show a mixed picture with around 15% of estimates agreeing with the model, approximately 37% are too high and around 48% are too low. Again, sample numbers are too low to enable conclusions to be made about the validity of the southern hemisphere sea ice uncertainty estimates. The EUSTACE uncertainty estimates for Tmax in the northern hemisphere MIZ, as is the case for Tmean, are good with around 76% of estimates in agreement with model values; the remaining 24% are too low.

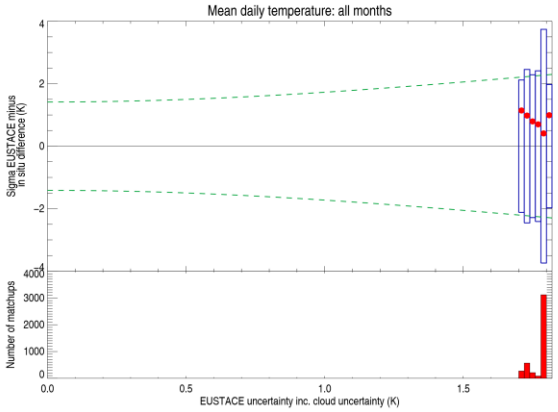
The analysis of EUSTACE Tmin uncertainty estimates (not shown) revealed EUSTACE estimates to be too low in all but a few cases.



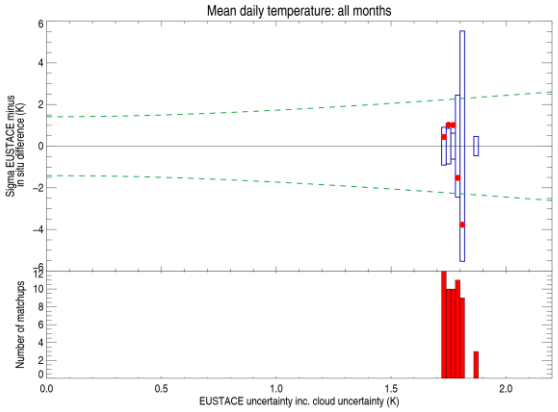
Land ice northern hemisphere



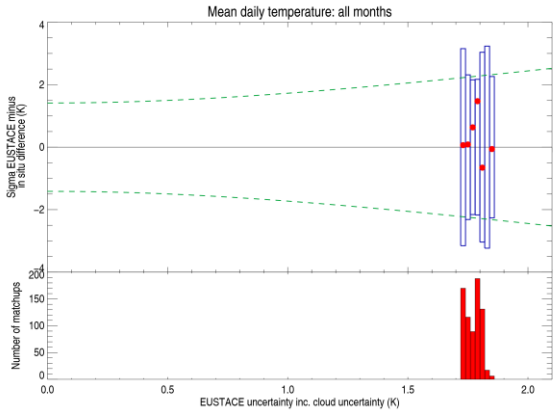
Land ice southern hemisphere



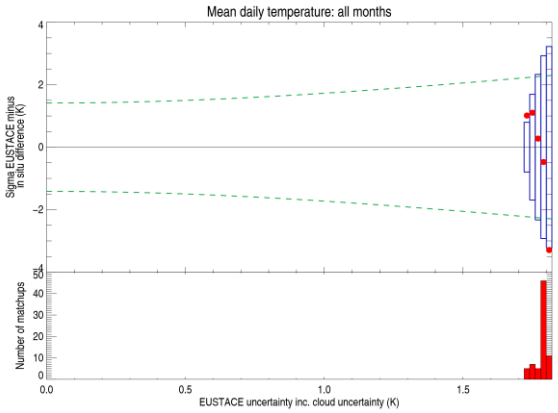
Sea ice northern hemisphere



Sea ice southern hemisphere

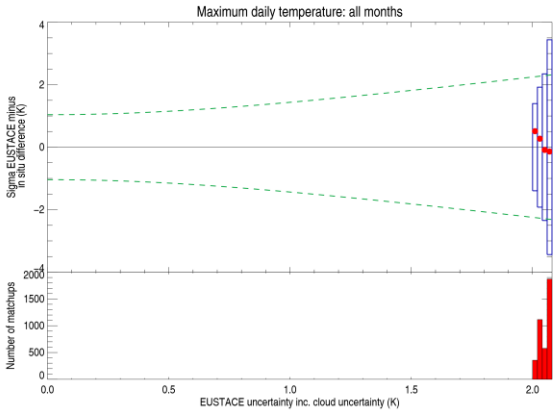


MIZ northern hemisphere

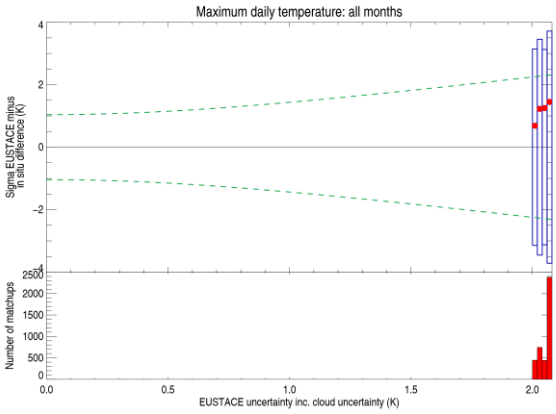


MIZ southern hemisphere

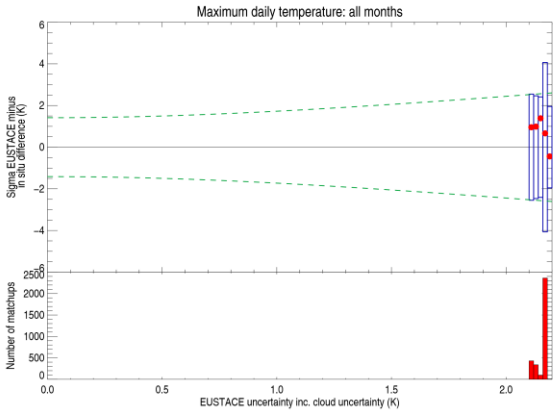
Figure 2-4: Validation of EUSTACE Ice Tmean uncertainties Green dashed lines show model estimates. Blue bars show match-up robust standard deviations. RSDs are always positive but are also plotted as negative bars to aid comparison with median values (red dots).



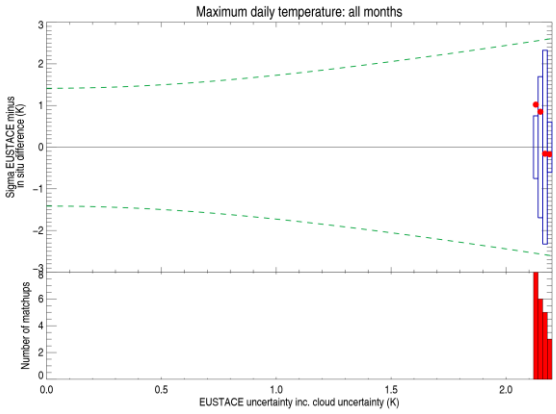
Land ice northern hemisphere



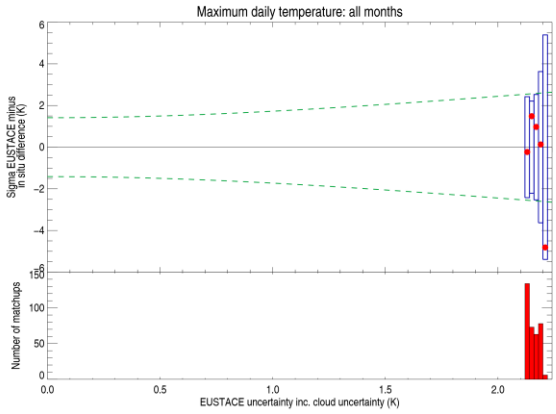
Land ice southern hemisphere



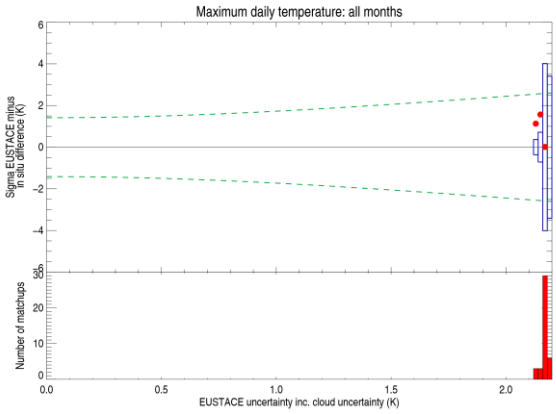
Sea ice northern hemisphere



Sea ice southern hemisphere



MIZ northern hemisphere



MIZ southern hemisphere

Figure 2-5: Validation of EUSTACE Ice Tmax uncertainties. See caption to **Figure 2-4** for explanation of plot components.

2.4. Summary

- EUSTACE test ice surface air temperature estimates and their uncertainty estimates have been compared with in situ air temperatures from automatic weather stations and sea ice buoys.
- Match-ups of EUSTACE ice mean daily temperaure estimates with in situ observations have a median discrepancy of -0.88 to +0.80 K depending on ice zone. Matchup RSDs are between 2.52 K and 3.43 K.
- Median discrepancy for EUSTACE ice minimum daily temperature estimates are between -1.39 K and 1.24 K, with RSDs between 3.39 K and 5.23 K depending on ice zones.
- Median discrepancy for EUSTACE ice maximum daily temperature estimates are between -1.20 K and 1.28 K, with RSDs between 1.49 K and 3.53 K depending on ice zones
- EUSTACE uncertainty estimates are generally too low. However, EUSTACE Tmean uncertainty estimates over northern hemisphere marginal sea ice agree with model estimates in around 55% of cases. Estimates of Tmax uncertainty over northern hemisphere land ice agree with the model in around 15% of cases and are too high in 37% of cases. Estimates of Tmax uncertainty over northern hemisphere marginal ice zone are good in 76% of cases
- Few match-ups to in situ data were found for southern hemisphere sea ice and marginal ice zones so that confidence in results for this region must be lower than in the other regions.

3.Validation of EUSTACE marine air temperature test data set

3.1. Data

EUSTACE marine air temperature test data (TD2) are provided in daily Level 3 files at a spatial resolution of 0.25° for 2001-2014. A single time stamp is provided for each file – no time information is provided at the pixel level; it is noted that time bounds attributes are not included in the files. In situ marine air temperature data are provided as point observations from HadNMAT2, the Hadley Centre Night Time Marine Air Temperature data set (Kent et al. 2013). One file is provided per day. Data have been quality controlled and corrected for changes in ship height (to 10m) as described in the paper. Both uncorrected and corrected data were provided but only the corrected data was used in this initial validation. HadNMAT2 data were used in the development of the regression between SST and air temperature, but only for the period 1940-2000. Data from 2001 onwards are therefore independent and can be drawn on for validation.

3.2. Approach

All daily in situ HadNMAT2 data files for each calendar year were collated into a single set of data. Subsequently, each daily EUSTACE test data file was read in turn and match-ups to HadNMAT2 identified. Yearly files were subsequently combined into a single file for analysis. For the purpose of generating the match-ups, each EUSTACE file is assigned a time of 12:00 noon on its local solar day of validity and a temporal match-up window of +/- 12 hours is used.

3.3. Results

3.3.1 Comparison of Test Dataset 1 and Test Dataset 2

Before looking at TD2 results in detail we first compare TD2 to Test Dataset 1 (TD1). A map showing the location of all match-ups between the EUSTACE TD1 and HadNMAT2 datasets for 2010 is shown in Figure 3-1. For direct comparison, the equivalent plot for TD2 is shown in Figure 3-2. Matches are shown as the difference between the two datasets and emphasise the spatial coverage of the HadNMAT2 data, which is dominated by the main shipping routes. Although some significant variation is evident in the region of the Gulf Stream, the results in Figure 3-1 do not offer much to aid understanding of any spatial or

temporal biases in the EUSTACE data. It is noted that the differences between EUSTACE and HadNMAT2 are warmer in TD2 compared to TD1.

Consequently, the match-ups were binned to 2° by 2° cells, and the resulting map is shown in Figure 3-3. Here some obvious spatial patterns exist with cool biases evident in the region of both the Gulf Stream and the Kuroshio. Other notable patterns in Figure 3-3 are cooler biases across most of the Pacific Ocean and warmer biases across the Indian Ocean. Also, notable positive biases are seen along the latitudinal extremes in both hemispheres. The equivalent plot for TD2 is shown in Figure 3-4, which also indicates that a warmer bias in TD2 compared to TD1

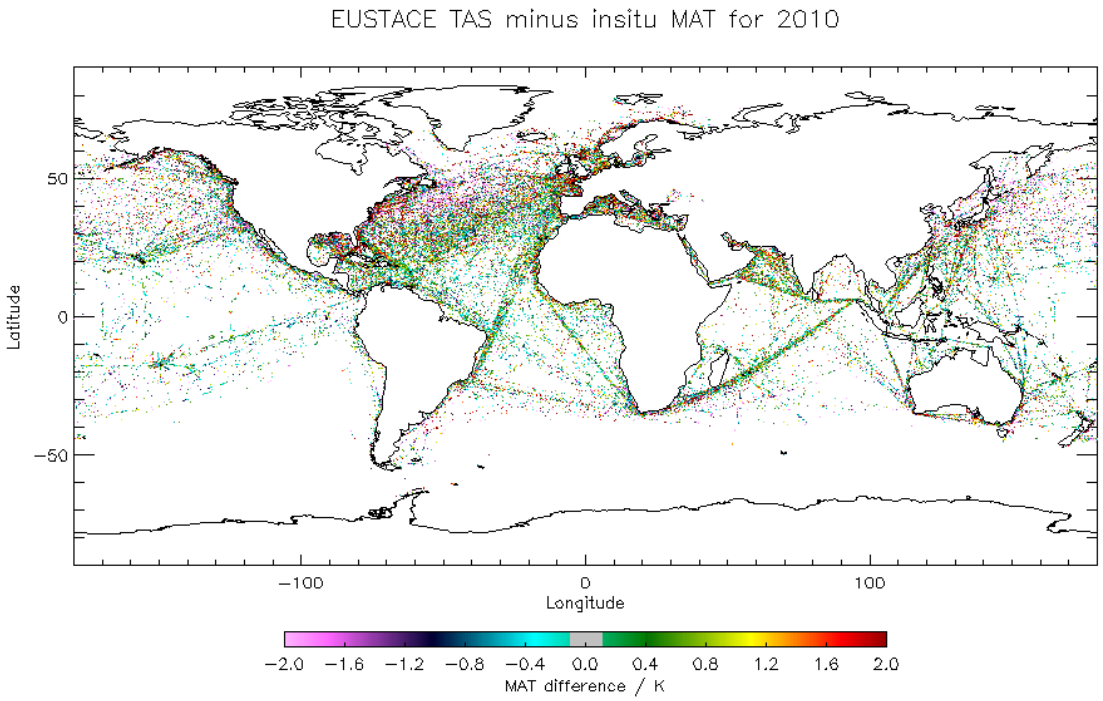


Figure 3-1: Location of all match-ups for 2010 between EUSTACE TD1 and HadNMAT2 marine air temperature. Data is plotted as the difference between the two datasets. The grey area indicates a range of +/- 0.1 °C

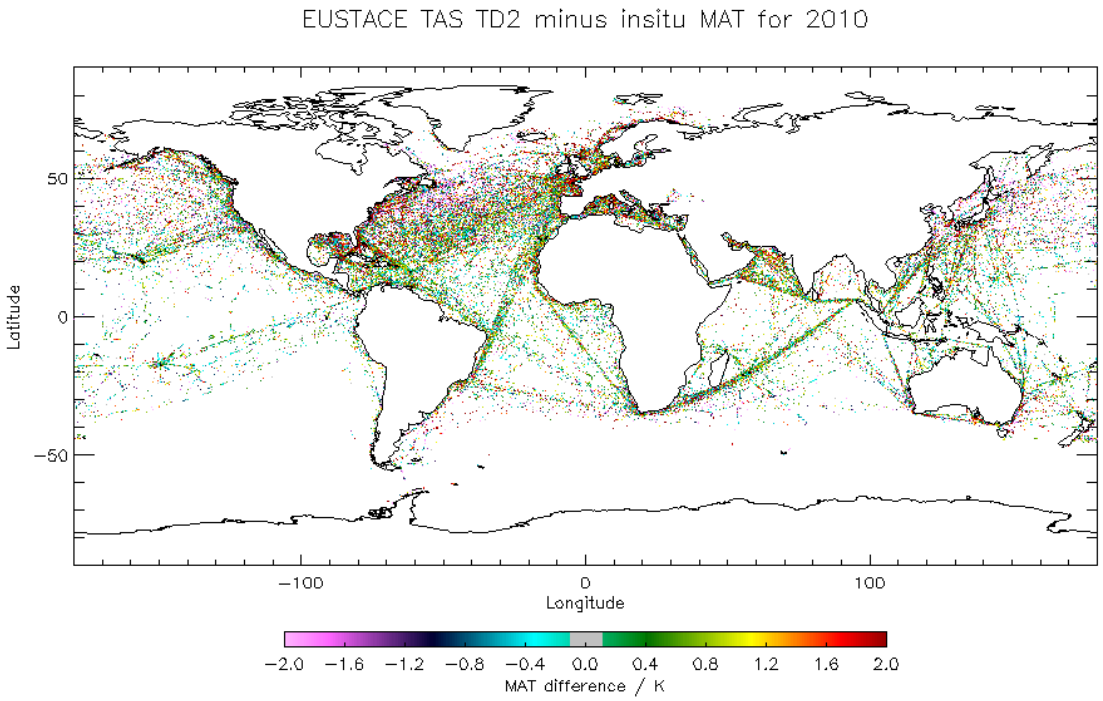


Figure 3-2: Location of all match-ups for 2010 between EUSTACE TD2 (V3) and HadNMAT2 marine air temperature. Data is plotted as the difference between the two datasets. The grey area indicates a range of +/- 0.1 °C

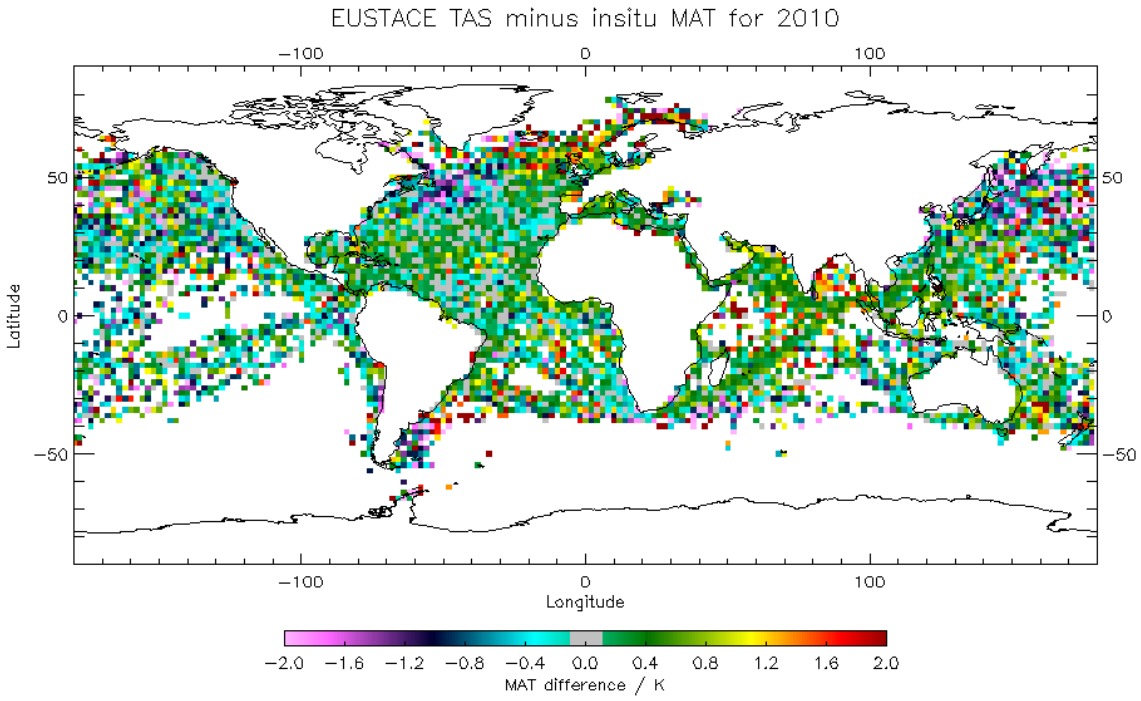


Figure 3-3: As Figure 3-1 but binned into 2° by 2° cells.

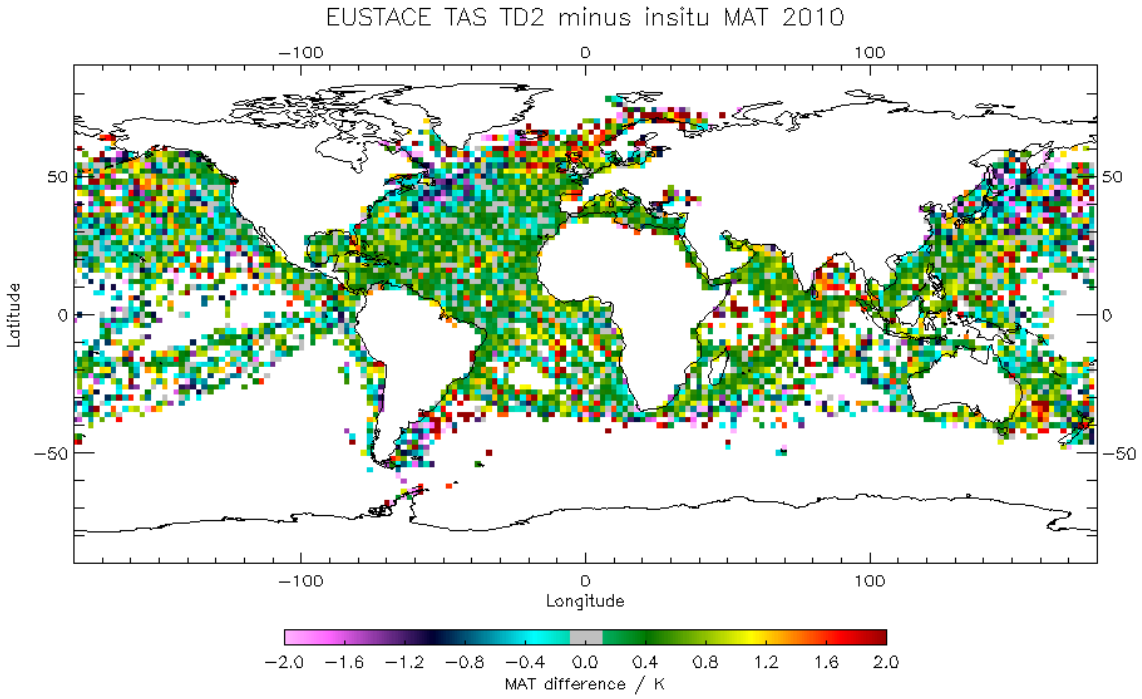


Figure 3-4: As Figure 3-1 but binned into 2° by 2° cells.

Normal and robust statistics of differences between EUSTACE TD1 and HadNMAT2 in situ marine air temperature for all of 2010 and for the seasonal periods of December, January and February (DJF), March, April and May (MAM), June, July and August (JJA) and September, October and November (SAM) are given in Table 3-1 for TD1 and in Table 3-2 for TD2.

In addition to warmer biases we observe notable changes in match-up numbers between the two data sets, particularly for the DJF period (19792 to 25129) and the SON period (26520 to 20098). While this change in apparent skill in TD2 should be explored further, the short period of the comparison (2010 only) may alias the year-to-year variability expected in the comparisons between the EUSTACE marine air temperature estimates and the reference data.

Table 3-1: Normal and robust statistics of differences between EUSTACE TD1 and HadNMAT2 in situ marine air temperature for all of 2010 and for the seasonal periods of December, January and February (DJF), March, April and May (MAM), June, July and August (JJA) and September, October and November (SAM).

	Number	Mean (°C)	SD (°C)	Median (°C)	RSD (°C)
All 2010	94065	+0.18	1.51	+0.14	1.20
DJF 2010	19792	+0.33	1.87	+0.23	1.39
MAM 2010	24523	+0.07	1.43	+0.04	1.14
JJA 2010	23230	+0.17	1.25	+0.16	1.08
SON 2010	26520	+0.20	1.48	+0.14	1.27

Table 3-2: Normal and robust statistics of differences between EUSTACE TD2 and HadNMAT2 in situ marine air temperature for all of 2010 and for the seasonal periods of December, January and February (DJF), March, April and May (MAM), June, July and August (JJA) and September, October and November (SAM).

	Number	Mean (°C)	SD (°C)	Median (°C)	RSD (°C)
All 2010	95707	+0.40	1.56	+0.34	1.21
DJF 2010	25129	+0.28	1.47	+0.26	1.14
MAM 2010	23560	+0.36	1.26	+0.35	1.08
JJA 2010	26920	+0.41	1.51	+0.34	1.28
SON 2010	20098	+0.59	1.97	+0.44	1.44

3.3.2 TD2 Results

This section contains results for TD2 only. We first look at the spatial variability of the differences between EUSTACE and HadNMAT2, we then consider the dependence of the results as a function of key variables. Finally in we look at the product uncertainties.

Spatial maps

Spatial maps showing the yearly variation of differences between EUSTACE TD2 and HadNMAT2 for 2002 through 2010 are shown in Figure 3-5. The coverage of match-ups is very consistent from year to year, as are the observed biases, which consistently show cold biases in the regions of the western boundary currents and warm biases along the latitudinal extremes.

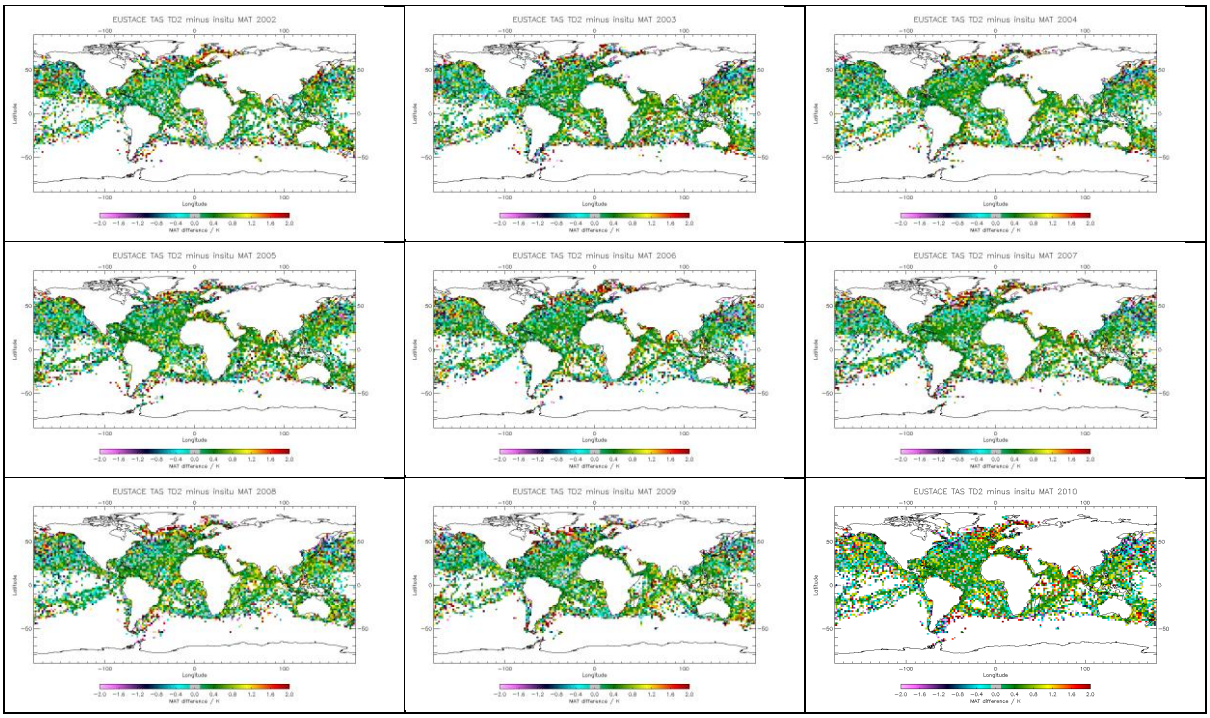


Figure 3-5: As Figure 3-1 but binned into 2° by 2° cells for (top row, left to right) 2002, 2003 and 2004, (middle row, left to right) 2005, 2006 and 2007, (bottom row, left to right) 2008, 2009 and 2010

Dependence Plots

The dependence of the difference between EUSTACE TD2 and HadNMAT2 as a function of latitude and date are shown in

Figure **3-6** and Figure 3-7, respectively. No notable trends are seen in either plot. The dependence on latitude shows significant changes at the two latitude extremes as seen in the spatial plots. However, the number of match-ups in these locations is very low, so the calculated differences when binned by latitude are heavily influenced by outliers.

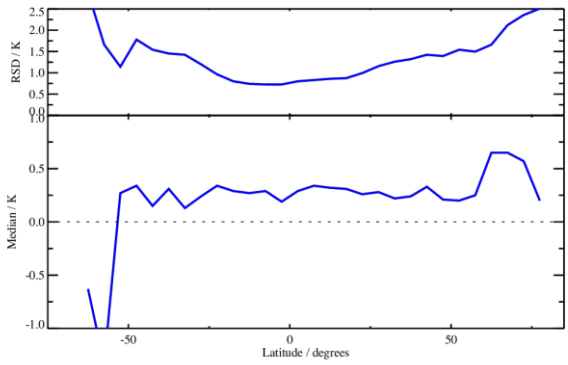


Figure 3-6: Difference between EUSTACE TD2 and HadNMAT2 as a function of latitude

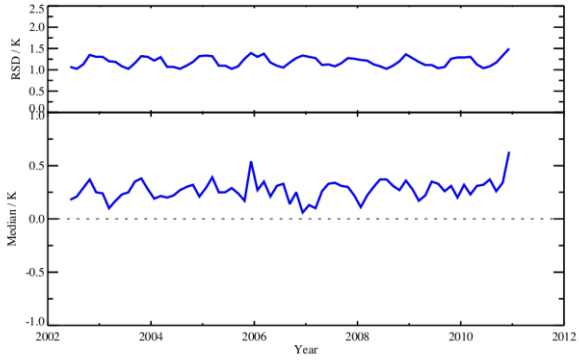


Figure 3-7: Difference between EUSTACE TD2 and HadNMAT2 as a function of date

One major difference between the EUSTACE TD2 and HadNMAT2 data is the time value assigned to each data point. The EUSTACE data are meant to represent the local solar daily mean at each location, whereas the HadNMAT2 data are measurements at a specific UTC time. To investigate if this difference influences the results, each HadNMAT2 measurement time was converted to a local solar time, and the dependence of the differences as a function of local solar time is shown in Figure 3-8. The consequence of using night time only data for validation is evident in Figure 3-8, as there are no match-ups for certain times of the day. This is demonstrated further in Figure 3-9, which shows the distribution of match-ups as a function of local solar time. It is arguable that the good range of values in Figure 3-8 is between 6 PM and 8 AM as a result.

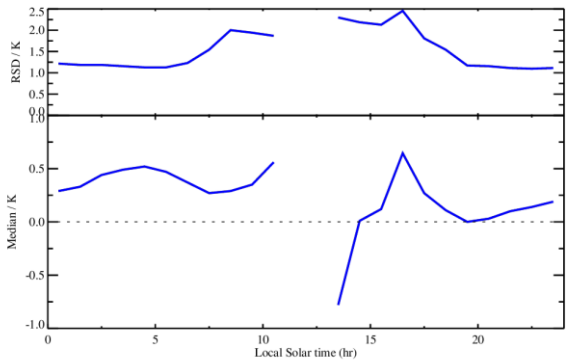


Figure 3-8: Difference between EUSTACE TD2 and HadNMAT2 as a function of local solar time

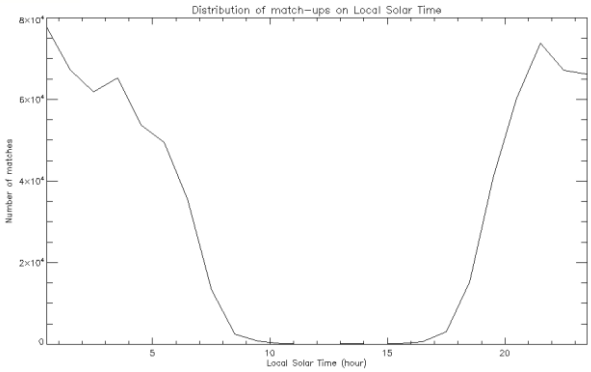


Figure 3-9: Distribution of match-ups as a function of local solar time

If we only look at the time period from 6 PM to 8 AM in Figure 3-8, then there is clear evidence of a dependence on local solar time with bias values ranging from +0.5 °C at around 5 AM to around 0.0 °C at 8 PM. To investigate the shape of this dependence, Figure 3-10 shows a synthetic daily cycle of air temperature (blue) along with its daily mean (red); the data in Figure 3-10 has been synthesised to show how a daily cycle of temperature, driven purely by solar input, compares to its daily mean.

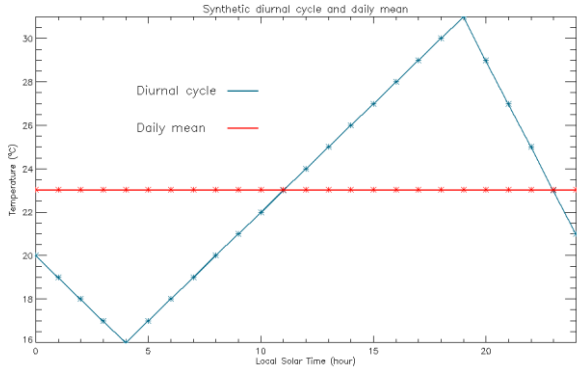


Figure 3-10: Synthetic diurnal cycle (blue) and daily mean (red) as a function of local solar time

The observed dependence on local solar time in Figure 3-8 is directly correlated to the difference between the daily cycle and its mean (red line minus blue line) in Figure 3-10 as expected for comparisons of data at different times of the day. We can also see in Figure 3-10 that the daily mean is equivalent to the daily cycle at two times of the day – roughly 11 AM and 11 PM. Although we do not have enough match-ups for assessing 11 AM, the observed bias for 11 PM in Figure 3-8 is close to the median bias presented in Table 3-2;

however, the residual difference suggests that the observed biases are not solely due to the differences in measurement time.

Histogram

The distribution of the observed differences between EUSTACE TD2 and HadNMAT2 is shown in Figure 3-11, which also shows the results from a linear least squares fit of a Gaussian Probability Distribution Function (PDF) to the data.

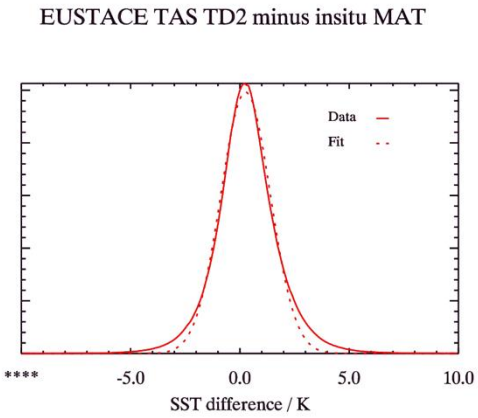


Figure 3-11: Distribution (solid line) of match-ups differences between EUSTACE TD2 and HadNMAT2. The dashed line represents the results form a linear least square fit of a Gaussian Probability Distribution Function to the data.

The results in Figure 3-11 show that the differences are well represented by the Gaussian PDF as the agreement between the data and the fit is very good with small variations in the wings of the distribution. No notable asymmetry is seen, indicating that residual cloud contamination in the input satellite SST data is not a major factor.

Uncertainty validation

To validate the uncertainties in the EUSTACE dataset we apply the same method as for land and ice. The principal approach to validation of uncertainties is to examine the distribution of EUSTACE minus HadNMAT2 differences as a function of uncertainty.

The results from applying this method to the EUSTACE TD2 uncertainties are shown in Figure 3-12. Here we have assumed the uncertainty of the in situ data is 1.0 K. The blue vertical lines span -1 to +1 standard deviation of discrepancy, for data binned into 2.0 K bins. When the EUSTACE uncertainty is small, the SD of discrepancy is dominated by the HadNMAT2 uncertainty; for larger EUSTACE uncertainties, the SD of discrepancy approaches the estimated uncertainty of the EUSTACE data. The green dotted line gives the locus of the results if the EUSTACE uncertainty is perfectly estimated and our uncertainty validation model and assumptions are valid. Deviations from the dotted line indicate biases in uncertainty estimation. The bias between EUSTACE and HadNMAT2 is also shown in Figure 3-12 by the red lines. The total uncertainty for each EUSTACE measurement was estimated by adding the random, locally correlated and correlated terms in quadrature.

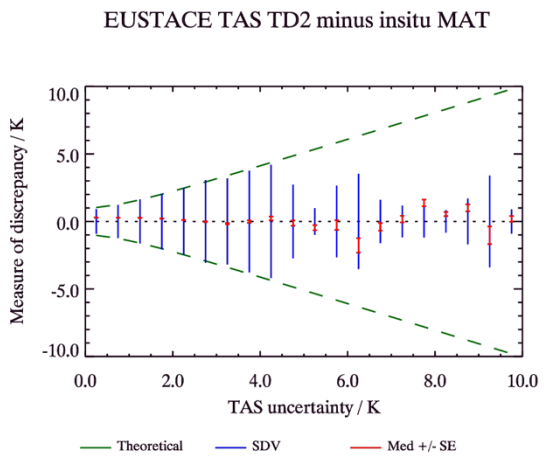


Figure 3-12: EUSTACE TD2 total uncertainty validation plot

The result in Figure 3-12 shows excellent agreement between the uncertainty validation model and the EUSTACE TD2 total uncertainty values for match-ups where the total uncertainty is below 4.5 K. For values above 4.5 K the uncertainties are significantly over estimated and highly variable, as the blue vertical lines do not match the green dotted line for uncertainties above 4.5 K.

Each of the five terms used to derive the total uncertainty was looked at individually to assess their contribution. The total uncertainty is dominated by one term – the correlated 2 term – which is the contribution that arises in the relationship building (estimating air temperature from SST) with a length scale of approximately 1500km. The magnitude and

variation of the locally correlated 2 term is shown in Figure 3-13, which is more or less identical to the total uncertainty shown in Figure 3-12.

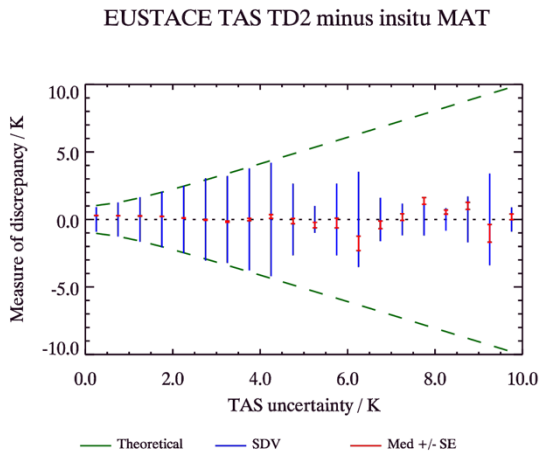


Figure 3-13: EUSTACE TD2 uncertainty validation – correlated uncertainty 2

The range of uncertainties in Figure 3-12 is very large, so Figure 3-14 shows the validation of the EUSTACE total uncertainty for 0.0 to 2.0 K at a resolution of 0.05 K. Figure 3-14 shows excellent discrimination between uncertainty values and excellent agreement between the uncertainty validation model and the EUSTACE uncertainties.

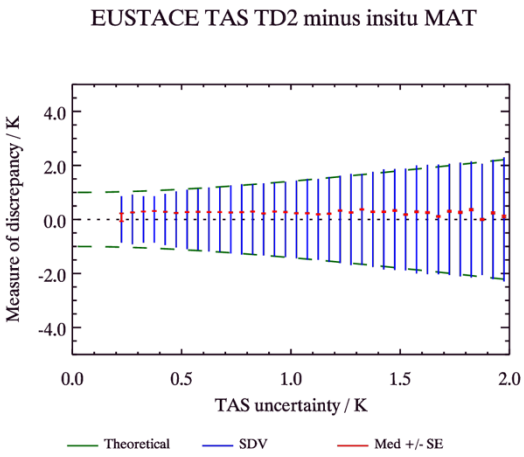


Figure 3-14: As Figure 3-12 but for a reduced range to emphasise the smaller uncertainty values.

The large overestimated uncertainties were looked at in a little more detail to look for any obvious patterns that may explain their distribution. Figure 3-15 show a time series of the

correlated uncertainty 2 term, which shows a dominant seasonal signal for the majority of match-ups, and a small number of outliers spread quite randomly across the time series.

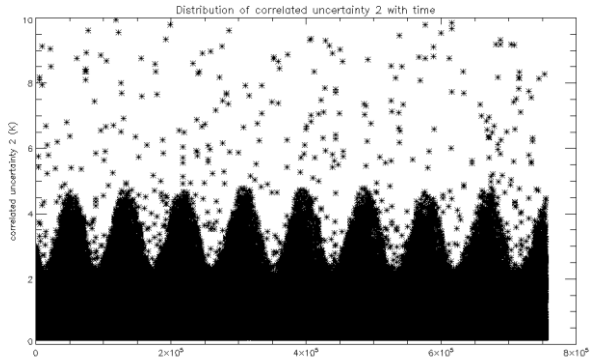


Figure 3-15: Distribution of correlated uncertainty 2 with time (match-ups are stored chronologically so plot covers May 2002 through December 2010).

Figure 3-16 shows the location of all match-ups where the total uncertainty is > 4.5 K. It is noticed that these often appear in the region of the Gulf Stream as well as the latitudinal extremes although their coverage is reasonably global.

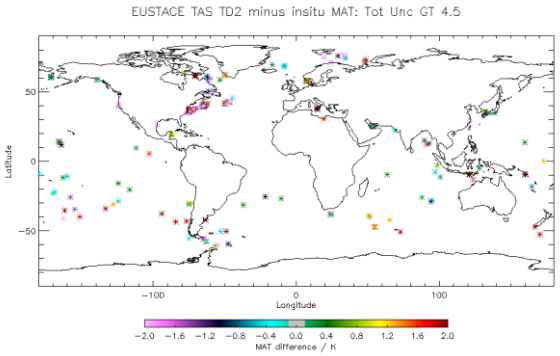


Figure 3-16: Location of matchups (462) where correlated uncertainty 2 is > 4.5 K.

3.4. Summary and conclusions

EUSTACE marine air temperature test data has been validated using HadNMAT2 in situ marine air temperature data for 2002 to 2010. In general, the EUSTACE data are warm biased compared to the in situ data. We conclude that:

- TD2 is biased warm (0.3-0.4 K) compared to TD1 (0.0-0.2 K)
    - Notable cold biases are observed over the major western boundary currents
    - Notable warm biases are observed at the northern and southern “edges” of the data
- The impact of validating the daily mean using night time only match-ups appears to be minimal
    - It does not account for the observed bias
- The EUSTACE uncertainty discrimination and magnitude is excellent up to values of around 4 K
    - The uncertainty is dominated by the correlated 2 term
    - A small number of matchups exist where the correlated 2 term has been significantly overestimated

4.References

ARM (2011), ARM Surface Meteorology Systems (MET) Handbook Section 7.0 Instrument details, ARM publication DOE/SC-ARM/TR-086, available from http://www.arm.gov/publications/tech_reports/handbooks/met_handbook.pdf with details at http://www.arm.gov/publications/tech_reports/handbooks/met/sgpmetdetaileddescription.pdf

C.E. Bulgin, O. Embury, G. Corlett and C.J. Merchant (2016). Independent uncertainty estimates for coefficient based sea surface temperature retrieval from the Along-Track Scanning Radiometer instruments, RSE, doi: 10.1016/j.rse.2016.02.022

D. Ghent, Land Surface Temperature Validation and Algorithm Verification (Report to European Space Agency). 2012(UL-NILU-ESA-LST-VAV).

E.C. Kent, N.A. Rayner, D.I. Berry, M. Saunby, B.I. Moat, J.J. Kennedy, D.E. Parker (2013). Global analysis of night marine air temperature and its uncertainty since 1880: the HadNMAT2 Dataset, JGR Atmos., doi: 10.1002/jgrd.50152