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 D*1.5*

Report on the relationship between satellite surface skin temperature and surface air temperature observations for oceans, land, sea ice and lakes

Deliverable Title	*Report on the relationship between satellite surface skin temperature and surface air temperature observations*
Brief Description	*This report will describe the derived relationships between satellite surface skin temperature observations and surface air temperature observations for land, lakes, ocean, sea ice and ice sheets (Task 1.4).*
WP number	1
Lead Beneficiary	*Jacob L. Høyer, Danish Meteorological Institute (DMI)*
Contributors	*Elizabeth Good (Met Office), Pia Nielsen-Englyst (DMI), Kristine Skovgaard Madsen (DMI), Iestyn Woolway (University of Reading), John Kennedy (Met Office)*
Creation Date	05/01/2017
Version Number	2
Version Date	05/04/2018
Deliverable Due Date Actual Delivery Date	M24 M40 (v2)
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	05/01/2017	Jacob Hoyer	Created
		Nick Rayner	Comments returned to authors
			Authors returned revised sections and submitted as draft
	21/12/2017	Lizzie Good	Revised Section 3.3 after reprocessing
2	05/04/2017	Nick Rayner	Submission version

Table of Contents

Acronym list

AASTI	Arctic and Antarctic Ice Surface Temperatures from thermal Infrared satellite sensors
AIRS	Atmospheric Infrared Sounder
ARM	Atmospheric Radiation Measurement programme
CCI	Climate Change Initiative
CMSAF	EUMETSAT's Climate and Monitoring, Satellite Application Facility
DEM	Digital Elevation Model
DMI	Danish Meteorological Institute
ERA	ECMWF Reanalysis
ESA	European Space Agency
EUSTACE	EU Surface Temperature for All Corners of Earth
FVC	Fraction of Vegetation Cover
GHCN-D	Global Historical Climate Network – Daily
HadISD	Hadley Centre Integrated Surface Daily
ICOADS	International Comprehensive Ocean-Atmosphere Data Set
IST	Ice Surface Temperatures
IAT	Near surface air temperatures over ice
JULES	Joint UK Land Environment Simulator
LSA-SAF	EUMETSAT's Land Surface Analysis Satellite Application Facility
LSAT	Land Surface Air Temperature
LST	Land Surface Temperature
LSWT	Lake Surface Water Temperature
LSTday	Daytime LST
LSTngt	Night time LST
MAD	Mean Absolute Difference
MAT	Marine Air Temperature

MODIS	MODerate resolution Imaging Spectroradiometer
NH	Northern Hemisphere
NMAT	Night time MAT
NASA	National Aeronautics and Space Administration
NWP	Numerical Weather Prediction
OSI-SAF	EUMETSAT's Ocean and Sea Ice Satellite Application Facility
PROMICE	Programme for Monitoring of the Greenland Ice Sheet
RMSD	Root mean square difference
SAT	surface air temperature
SEVIRI	Spinning Enhanced Visible and InfraRed Imager
SH	Southern Hemisphere
SST	Sea Surface Temperature
SZA	Solar zenith angle
T2m	Screen-level air temperature
Tavg	Averaged daily T2m
Tmax	Maximum daily T2m
Tmin	Minimum daily T2m
Tskin	Skin temperature
UoL	University of Leicester
WP1	EUSTACE Work package 1
WP2	EUSTACE Work package 2
WP3	EUSTACE Work package 3

1. Executive Summary

This report describes work to develop an understanding of relationships between satellite surface temperature observations and surface air temperature, in order to use satellite retrievals to derive air temperature fields with unprecedented detail and accuracy and thus increase the density of surface air temperature information globally. The different surfaces addressed comprise ocean, land, ice and lakes. Empirical relationships have been determined using measurements made in situ, satellite observations and auxiliary information. The relationships have been determined separately for each surface type to account for the different physical conditions present. The land and ice regions use the satellite skin observations directly in combination with auxiliary information, whereas the ocean approach applies a spatially and seasonally varying offset to the observed sea surface temperatures. The final results, however, can be used to estimate air temperature on the same global grid for all surfaces with the same definition of a day.

To estimate Marine Air Temperatures (MAT), the offset was calculated from Sea Surface Temperature (SST) and MAT measurements made by ships between 1963 and 2000. Validation against independent observations shows biases in the satellite derived MAT of around 0.1 to 0.2°C. See section 3.2 for details on the MAT results.

For the land, described in section 3.3, the main challenges for estimating Land Surface Air Temperature (LSAT) from satellite observed Land Surface Temperatures (LST), are the heterogeneity of the surfaces and residual cloud effects. The derived LSAT gives median differences within +/- 0.1°C and root mean square differences (RMSD) between 2.80 and 4.89°C, when compared against more than 10,000 independent observations. Estimation of the daily minimum temperature appears to be more accurate than daily maximum estimation.

The high latitude ice-covered regions, presented in section 3.4, are challenging due to residual cloud contamination in the satellite observations and data sparsity. A data rescue effort was needed to secure in situ observations over the Arctic Sea Ice, Greenland Ice Sheet, Southern Ocean Sea Ice and the Antarctic Ice Sheet. The mean daily Ice surface Air Temperatures (IAT) derived from satellite observations have biases below 0.4°C and RMSD between 3.11 and 3.55°C for the four ice covered regions. The performance in estimating IAT is similar to that of ERA-Interim for all regions, except for the Greenland Ice Sheet where the satellite estimated IAT are consistently better.

A different methodology had to be chosen over the lakes due to the lack of in situ observations there, as discussed in section 3.5. A modelling approach has been followed to assess the influence of lakes on LSAT to permit subsequent EUSTACE analyses to be more accurate in affected regions when meteorological observations are scarce. In addition, a method for estimating Lake Surface Water Temperature (LSWT) from nearby air temperature observations has been demonstrated to generate accurate time series of LSWTs.

Simple tests using a physical model to estimate air temperature from satellite retrievals over ocean, land and ice have shown promising results and may be able to provide alternative relationship models (section 3.6). It is shown, however, that the estimated surface air temperatures are tightly coupled to the Numerical Weather Prediction (NWP) model fields used as input. This will make the performance dependent upon the quality of the NWP and will only to a limited degree include new information from the satellite observations.

In general, the outcome of the work within EUSTACE Task 1.4 "Derive relationships between different observation types for land, lakes, ocean, sea ice, and ice sheets" is very encouraging. Innovative methods have been developed that facilitate the use of satellite observations to provide new and unique information on global surface air temperature.

However, estimation of surface air temperature from satellite observations is not a "done deal" when EUSTACE is finished. The lessons learned and future work in section 4 outline some of the difficulties that were met and highlight areas where future work could bring improvements in near surface air temperature estimates. Key areas include improvements in: surface air temperature consistency where two surface types meet (edgelands), deriving and understanding uncertainty estimates, inclusion of more satellite and in situ observations and a further utilization of the physical models.

2. Project Objectives

With this deliverable, the project has contributed to the achievement of the following objectives (DOA, Section B1.1):

No.	Objective	Yes	No
1	Intensively develop the hitherto immature use of Earth Observation estimates of Earth's surface **skin** temperature to enable new Climate Data Records of the surface **air** temperature Essential Climate Variable (ECV) to be created, for all locations over all surfaces of Earth (i.e. land, ocean, ice and lakes), for every day since 1850. EUSTACE will achieve this by: combining information estimated from multiple satellites with surface air temperature measurements made *in situ* and **creating complete analyses** of surface air temperature, through the application of novel statistical in-filling methods.	x	
2	Integrate these new daily surface air temperature Climate Data Records into a range of applications in Earth System Science and Climate Services and research, amongst others. EUSTACE will achieve this via the active and continuous engagement of trail-blazer users, and the provision of products through already-existing user community data portals and service mechanisms, in standard formats.		x
3	Undertake and report detailed research into the relationships between surface skin temperature estimated from Earth Observation satellite measurements and surface air temperature observed *in situ* by conventional measurements, over all surfaces of the Earth, including the polar regions. This is likely to provide information useful for refining coupling in Earth system models.	x	

4	Create a sustainable, automated system at an appropriate level of maturity for the potential production of the products beyond the lifetime of the project. To enable this, EUSTACE will also identify Earth Observation and conventional data streams that could be used to update the surface air temperature Climate Data Records in the future, including those from Sentinel missions.		x
5	Extensively validate the new surface air temperature Climate Data Records against independent, surface-based reference data, sourced by the project for this purpose.	x	
6	Develop and report new, consistent, validated estimates of uncertainty both in already-existing Earth Observation surface skin temperature estimates and in the new surface air temperature Climate Data Records, at all locations and times across the Earth's surface.	x	
7	Develop links with related activities within Europe and beyond to help to ensure the execution of a joined-up work programme, the Copernicus Services and to enable the provision of requirements for the future surface skin temperature and surface air temperature observing system.		x
8	Other – not directly linked to one of the above objectives		x

3. Detailed Report

3.1 General Introduction

Daily near surface air temperature is one of the most crucial variables used in weather and climate science and service provision around the world and is one of the Global Climate Observing System (GCOS) Essential Climate Variables (ECV). Traditionally, information on near surface air temperature has been obtained from meteorological station measurements. However, air temperature observations are not available everywhere on Earth and vast areas exist with very limited observations e.g., the Polar Regions and Southern Ocean. Moreover, the time series available may contain gaps or have limited length and the observations at meteorological stations may not be representative of air temperatures at neighbouring locations.

This work has been carried out to develop methods to estimate near surface air temperatures from satellite data for all surfaces of the Earth: oceans, land, lakes and ice, to supplement the available in situ observations. To achieve this, detailed research into the relationship between air and surface (skin) temperatures has been carried out. Temporally and spatially collocated observations have been used to investigate their relationship in well-observed regions and afterwards the findings have been applied to estimate air temperatures for all surfaces of the Earth.

Due to physical differences and differences in the observed heat fluxes and temperature variability, the skin to near air surface relationships have been explored separately for each of the surfaces: oceans, land, lakes and ice. However, a close communication and exchange of ideas has been maintained between the EUSTACE partners, to ensure that consistent products have been developed, to facilitate user uptake and to aim for a consistent product at the boundaries between the products, such as the land-ocean boundary or the ocean-sea ice interface.

This report describes the final outcome of EUSTACE WP1 Task 1.4, where relationships between satellite-observed surface temperatures and near surface air temperatures have been developed. The relationships and the derived near surface air temperatures are important stand-alone products; they will also be used in the EUSTACE WP2: "Data set construction".

Sections 3.2-3.4 report the work carried out independently for each surface to develop the empirical models for the ocean, ice and land regions. To help the reader, we have aimed at the same structure for the each of these sections. Due to data sparseness, the empirical approach could not be followed for lakes. Section 3.5 therefore describes an alternative approach for lakes that deviates from that employed for the other surfaces.

Results from a collaborative effort using a simple physical model to derive near surface air temperatures from satellite skin temperature observations over oceans, land and ice are described in section 3.6. This section serves as an outlook to assess the potential of the physical models in next generation products.

3.2 Relationships over oceans

3.2.1 Summary of ocean relationships

Marine air temperatures (MAT) were estimated from sea-surface temperature (SST) both measured in situ and retrieved from satellite measurements. MAT was calculated by adding a spatially and seasonally-varying offset to the SST. The offset was calculated using bias-adjusted SST and MAT measurements made by ships between 1963 and 2000. Uncertainty in the estimated MAT was estimated by analysing the variance of daily deviations from the climatological average offset. The average offsets were then used to predict MAT from SST for ships between 1995 and 2005. This overlaps the training period to test if there is a discontinuity at the ends of the training period. The mean bias was around 0.1 to -0.2°C, and the normalised standard deviation of errors was between 0.7 and 1.3, which indicates that the uncertainties are correct to within 30%. The errors were normalised by the standard deviation because the standard deviations (and the estimated range of the errors) vary between 0.1°C and 5.0°C. Application to satellite data and comparison of the satellite-derived air temperatures to in situ data not used in the estimation of the relationships shows a small warm bias of around 0.1-0.2°C, which indicates there is a residual

systematic error of order 0.1°C, which is consistent with the estimated uncertainty associated with systematic errors.

3.2.2 Data

Assessment of the sea-surface temperature/air temperature relationship over the oceans was based on analysis of in situ data from ICOADS release 2.5.1 (Woodruff et al. 2011).

The data are taken from ship reports in which both sea-surface temperature (SST) and marine air temperature (MAT) were measured. During the day, MAT measurements from ships are biased warm due to solar heating of the ships' superstructures so only night time MAT (NMAT) observations are used here. Night time is defined to be the period between one hour after sunset and one hour after sunrise to allow time for the effect of the ships' heating to dissipate.

Solar heating is not the only source of bias in the NMAT measurements. Changes in the height of the observing deck of the ship can also lead to artificial trends in the data if left uncorrected. Adjustments were applied to the NMAT data to account for changes in the measurement height. The adjustments were based on the method described in Kent et al. (2013), but referenced to a consistent observing height of 2m rather than 10m for consistency with the other surface types in EUSTACE.

SST measurements also suffer from systematic errors. The sources of these errors and the adjustments that are applied to account for them are described in Kennedy et al. (2011). Adjustments developed for individual SST measurements were applied to the SST reports as described in Atkinson et al. (2013). The adjustments aim to ensure that the SSTs are consistent with drifting buoy data, which have a nominal measurement depth of 0.2m.

The NMAT and SST data were quality controlled (QCed) using a system based on an update of Rayner et al. (2006) (Kennedy et al. in prep). 10% of the observations in each year were withheld for validation in WP3.

Additional in situ data for estimating the diurnal temperature range were sourced from the Global Tropical Moored Buoy Array and from research moorings archived at Woods Hole Oceanographic Institute (http://uop.whoi.edu/projects/projects.html). These moored buoys make frequent measurements of air temperature and near-surface water temperature. For some of the moorings however, the air temperature sensors suffer from ventilation errors, which affect day time observations (Anderson and Baumgartner 1998).

The modelled relationships derived from the in situ data were used to estimate air temperatures based on satellite-retrieved SSTs. The satellite-retrieved SSTs were a version of the SST CCI ATSR dataset (Merchant et al. 2014) regridded to daily 0.25° latitude/longitude resolution. This data set is based on the Along Track Scanning Radiometer (ATSR) series of instruments and is provided with uncertainty information, which is decomposed into three components associated with errors arising from: uncorrelated effects, locally correlated effects and systematic effects. The locally correlated error component is assumed to have a temporal scale of around 1 day and a length scale of 100km. The systematic error is assumed to be correlated perfectly in space and time, but is different for each of the satellite sensors of which two were used here: ATSR2 and AATSR. ATSR1 was not used owing to remaining difficulties with retrievals from the ATSR1 instrument.

3.2.3 Variability of the SST vs MAT relationship and our model

Existing data sets of air-sea temperature difference (Parker et al. 1995) show that the air temperature is, on average, 1 to 2°C cooler than the sea-surface but with significant geographical and seasonal variations. In some areas, such as at the continental margins, there are significant deviations from the climatological average. Therefore, the modelled relationships between sea-surface temperature and air temperature developed here are based on an assessment of the climatological average difference (used to infer MAT from SST) and on the climatological variance of the difference (used to estimate the uncertainty in the MAT).

Given an estimate of SST the MAT is modelled as:

$$MAT(s,t) = SST(s,t) + \delta(s,t) + error \qquad (Eq. 3.2.1)$$

where δ is the climatological average offset modelled as a Fourier series,

$$\delta(s,t) = a0(s) + a1(s)Sin(2\pi d/365) + a2(s)Cos(2\pi d/365) + a3(s)Sin(4\pi d/365) + a4(s)Cos(4\pi d/365)$$

$$(Eq\ 3.2.2)$$

where d is the day of the year (from 0 to 365), the a_i are the Fourier coefficients estimated without interpolation, and the error term has components arising from:

1. SST measurement error which, in the case of satellite data, will itself have uncorrelated, locally correlated and systematic components.
2. The climatological variance of the air-sea temperature difference, which is assumed to be normally distributed with mean zero and variance $\sigma var2(s,d)$ (estimation of the variance is described in Section 3.2.3.3) and is locally correlated in space and time (length and time scales are described in Section 3.2.3.5).

$$\sigma var2(s,d) = b0(s) + b1(s)Sin(2\pi d/365) + b2(s)Cos(2\pi d/365) + b3(s)Sin(4\pi d/365) + b4(s)Cos(4\pi d/365)$$

3. Systematic uncertainty in the bias adjustments applied to the in situ data, estimated to be around 0.1°C and correlated in space and time.
4. The uncertainty in the Fourier coefficients, a0, a1, a2, a3 and a4, are described by the error covariance matrix cov(ai, aj) and the square roots of the diagonal terms are given in the netcdf file as. The errors in the Fourier components are assumed to be uncorrelated in space.

NMAT-SST differences were calculated for each marine report in ICOADS v2.5.1 that had both an SST and NMAT measurement that passed QC. These Air-Sea Temperature differences (AST) were averaged onto a regular 1° latitude by 1° longitude by 5-day grid from 1963 to 2000.

3.2.3.1 Calculating a climatology of mean air-sea temperature difference

A 1° by 1° by 5-day AST climatology was calculated by averaging together all equivalent 5-day periods from 1963 to 2000 that contained data. A mean offset and the coefficients of four Fourier components were then estimated using multiple linear regression to obtain a smoothed annual cycle for that grid cell. The four Fourier components are:

$$Sin(2\pi d/365), Cos(2\pi d/365), Sin(4\pi d/365) \text{ and } Cos(4\pi d/365)$$

Where d is the day in the year starting from 0 on January 1st and ending with 364 on December 31st. This allows us to estimate an annual cycle in a grid box, even if there is not a complete 365 days' data there when averaged over 1963-2000.

The Fourier coefficients estimated at 1° by 1° resolution contained some outliers. Removing grid cells where the coefficients exceeded 300 or the estimated uncertainty in the coefficient exceeded 180 removed most obvious outliers.

Uncertainties in the weightings given to each of the Fourier components were estimated from the multiple regressions and used to estimate the uncertainty in the climatology. By construction, the errors in each coefficient lead to errors in the estimated air temperatures that are fully correlated in time, but negligibly correlated in space.

The climatological fields derived from these parameters remain somewhat noisy and incomplete (Figure 3.2.1). In some areas there have been too few coincident NMAT-SST observations in the period 1963 to 2000 to obtain a reliable estimate even with the Fourier interpolation.

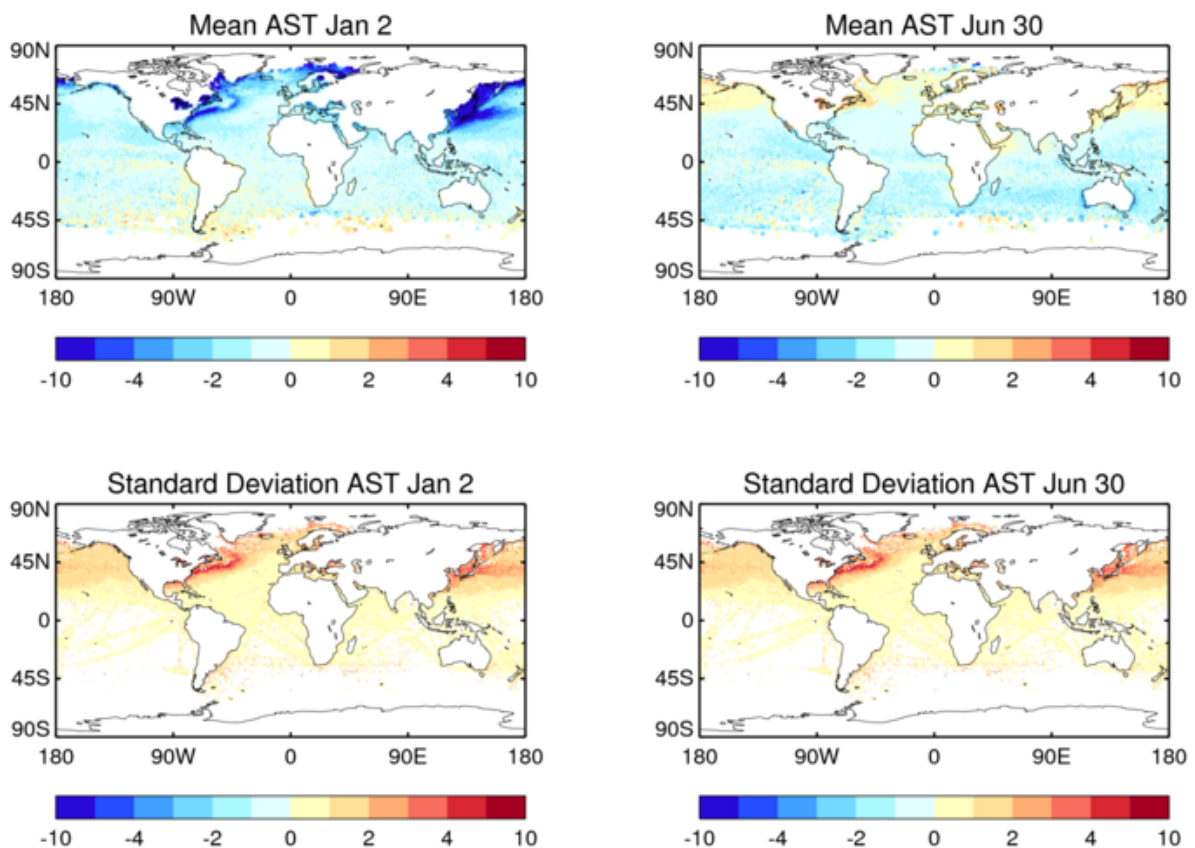*Figure 3.2.1:* *Mean air-sea temperature difference (°C) from the Fourier-smoothed climatology (top) for January 2nd and June 30th. Standard deviation of the air-sea temperature difference (°C) from the Fourier-smoothed climatology (bottom) for January 2nd and June 30th.*

3.2.3.2 Interpolating the climatology of the mean air-sea temperature difference

A second climatology, with improved coverage, was produced by interpolating the Fourier coefficients using Kriging.

First, a separate 5° by 5° by 1-month AST climatology was calculated by averaging the 1° by 1° by 5-day climatology. Coefficients for the same four Fourier components were then estimated based on the monthly data. The coefficients estimated at 5° by 5° resolution also contained a small number of outliers close to the ice edge which were removed by removing cells where the coefficients exceeded 500 or the uncertainty was greater than 110.

The coverage of coefficients at both 1° (from Section 3.2.3.1) and 5° resolution was then expanded slightly using a neighbour average and the 5° coefficients were interpolated to 1° using bilinear interpolation. The coverage of the interpolated 5° coefficients was near global.

For each Fourier component, differences were calculated between the coefficients originally estimated at 1° and the 1° estimates interpolated from 5°. This residual was then Kriged using fixed length and scale parameters and the Kriged residuals added back to recover a near-complete, smoothed estimate of the coefficients (Figure 3.2.2).

Uncertainties in the coefficients of each of the Fourier components were estimated from the multiple regressions (at 5° resolution) and used to estimate the uncertainty in the climatology. By construction, the uncertainties in the coefficients are perfectly correlated in time, but after Kriging, they are also somewhat correlated in space too in a way that is not easy to summarise or estimate, so it is recommended that the Kriged values are not used for infilling. The resulting fields are generally smoother, except at the edge of the observed regions where uncertainties are locally large.
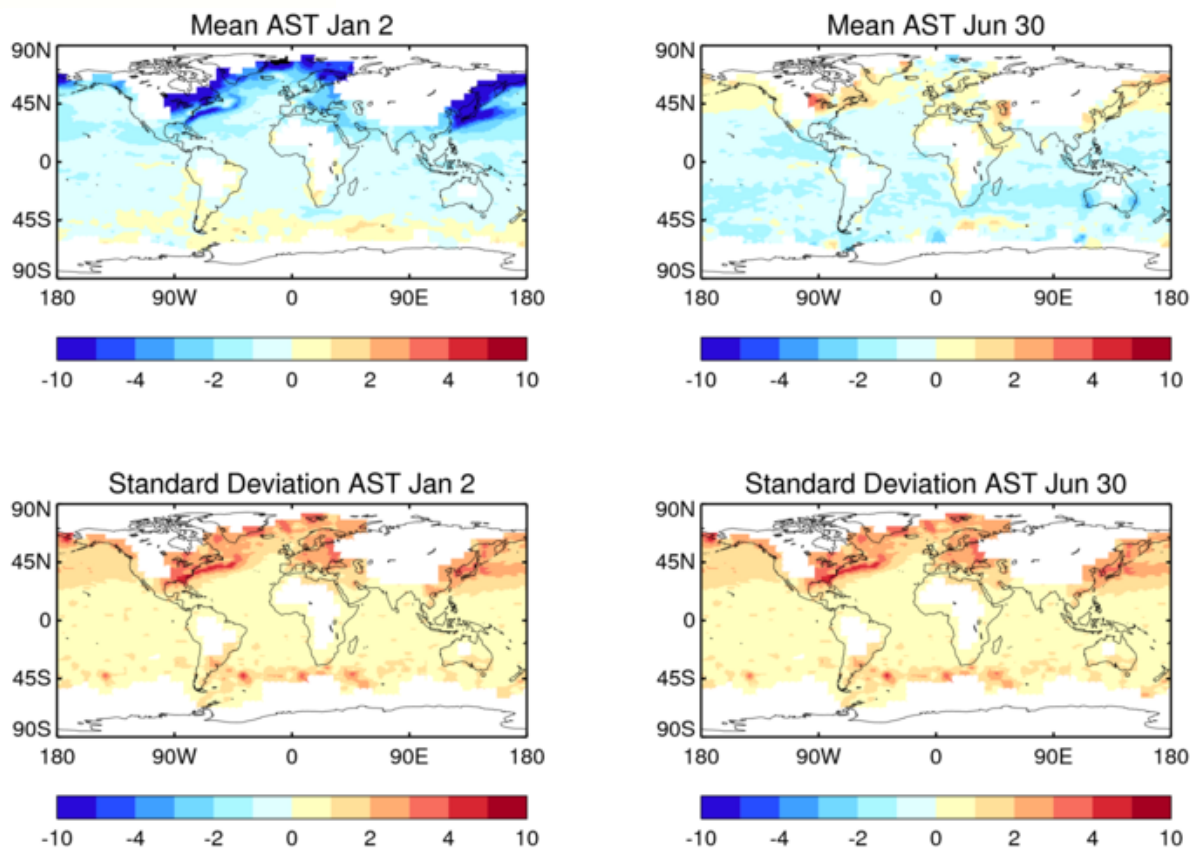
Figure 3.2.2: *Mean air-sea temperature difference (°C) from the Fourier-smoothed and Kriged climatology (top) for January 2nd and June 30th. Standard deviation of the air-sea temperature difference (°C) from the Fourier-smoothed and Kriged climatology (bottom) for January 2nd and June 30th.*

3.2.3.3 Calculating a climatology of air-sea temperature difference variability – estimating uncertainty in the air temperature estimates

Daily residuals from the Fourier-smoothed climatology (the non-Kriged version, initially, see section 3.2.3.4) were calculated for every day from January 1st 1963 to December 31st 2000 and averaged on to a regular 1° latitude by 1° longitude by 1-day grid. Note this is different from the 5-day grid used to calculate the mean as we are interested particularly in the daily variability for EUSTACE. Analysing the daily differences between the climatological AST and the actual AST will provide an estimate of the uncertainty in the air temperatures derived using the relationships and the spatial and temporal correlations of the likely errors (Section 3.2.3.5).

A 1° by 1° by 1-day AST standard deviation climatology was calculated for each day using data in a 9-day window centred on the day of interest. The AST standard deviation was calculated by assuming that deviations from the climatology were normally distributed with mean zero and a variance equal to the sum of the squares of the AST standard deviation and the measurement error uncertainty expressed as a standard deviation. The measurement error uncertainty for the daily grid cell averages was estimated to be approximately 1.4°C divided by the square root of the number of observations that contributed to the daily average. The AST standard deviation was varied between zero and 5°C in 0.05°C increments and the chosen value was the one that

maximised the likelihood of the data assuming the deviations were normally distributed. The uncertainty of 1.4°C assumes that both the SST and MAT measurements from the ship each had a measurement error uncertainty of around 1°C (Kent and Berry 2008). In some grid cells, the best estimate standard deviation was zero, which is unlikely to be correct. A lower bound of 0.3°C was set, consistent with the lower range of estimated AST variability in well-observed regions.

A mean and four Fourier components were then fitted to the estimated daily AST standard deviations for each 1° by 1° grid cell to obtain a smoothed annual cycle for the standard deviation in that grid cell (Figure 3.2.1).

Grid cells where the coefficients exceeded 2000 or the uncertainty exceeded 10000 were set to missing. These thresholds were determined by a visual inspection of the data. As before a monthly 5° data set was produced and the coefficients were re-estimated at this scale and then interpolated back to 1°. Grid cells for which the coefficients estimated from the daily data were more than 100 below those estimated from the monthly data and where the uncertainty was unbelievably small (less than 2) were set to missing. Grid cells for which the coefficients estimated from the daily data were more than 200 below those estimated from the monthly data were set to missing as long as they were not in the Gulf Stream. In other high gradient regions, this was not a problem.

As with the mean fields, the climatology produced without interpolation gives rather poor coverage in some areas.

3.2.3.4 Interpolating the climatology of air-sea temperature difference variability

Daily residuals from the Fourier-smoothed climatology (the Kriged version this time) were calculated for every day from January 1st 1963 to December 31st 2000 and averaged on to a regular 1° latitude by 1° longitude by 1-day grid. The procedure then continued as above. As in Section 3.2.3.2, the residual differences between the daily and monthly estimated coefficients were Kriged and then Kriged residuals were added back onto the monthly-estimated coefficients. Results are shown in Figure 3.2.2.

3.2.3.5 Estimating the temporal and length scales of air-sea temperature difference variability

A correlation length scale for the AST standard deviation (and hence the errors in the estimated air temperature) was estimated by calculating a variogram from the gridded daily data. The root mean-squared difference between AST anomalies as a function of separation between the two measurements (Figure 3.2.3) increases steadily from zero km to around 2500km, suggesting an exponential decay in correlation with an approximate length scale (e-folding) of around 1000km. This is a global estimate. Likewise, a temporal variogram was calculated by taking the root mean-squared difference between AST anomalies as a function of separation between the two measurements in time. The variance increases rapidly at first and then more slowly. The initial rapid rise is likely due to measurement error in the estimates and the slow rise more likely arises from correlations in the air temperature field itself, implying a temporal scale for the uncertainty in the air temperature estimated of between 3 and 4 days.
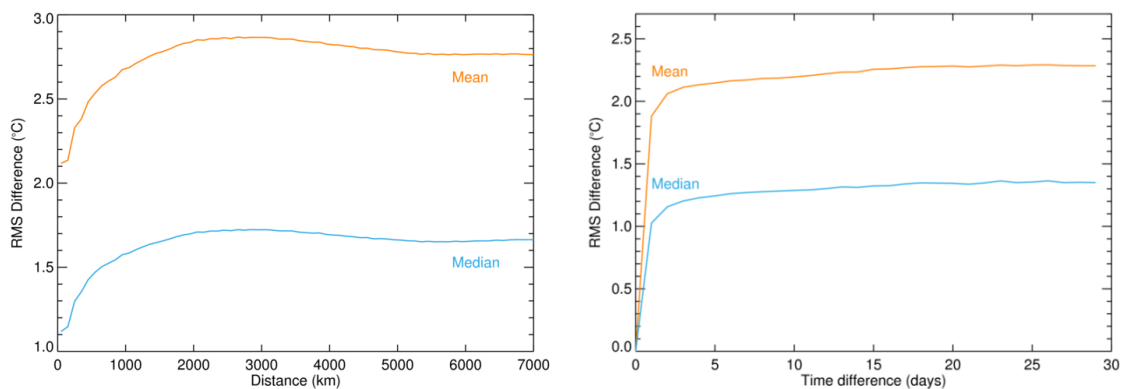
Figure 3.2.3: estimated spatial (left) and temporal (right) variograms showing the root mean (and median) squared difference between daily gridded air temperature measurements as a function of separation in space (in km) and time (in days).

3.2.3.6 Diurnal variability

The relationships developed here are intended to provide an estimate of the air temperature at a time which is consistent with the sea-surface temperature retrieval. If the input SST was measured at 10.30am (as the satellite SSTs used in EUSTACE are) then the output marine air temperature should be the air temperature as measured at 10.30am. Any differences in the relationship through the daily cycle should therefore be absorbed into the estimated uncertainty.

However, for the purposes of EUSTACE (where we are exploring the feasibility of providing information on Tmax and Tmin, as well as Tmean), it would be helpful to have a preliminary estimate of the diurnal temperature range (DTR) over the ocean. A low-resolution climatology of DTR in SST exists, based on monthly averages of drifting buoy data (Kennedy et al. 2007); no such data set exists for MAT.

To assess the relationship between the average SST DTR and MAT DTR, data were obtained from research moorings. A number of these, including the Global Tropical Moored Buoy Array, are affected by solar heating biases which tend to increase the DTR of measured MAT. Data from research moorings provided by Woods Hole Oceanographic Institute were used to calculate average diurnal cycles for each mooring making measurements of both air temperature and sea-surface temperature.

For those moorings which do not have an obvious daytime heating bias, the average DTR for MAT was larger by 0.2 to 0.5°C than the average DTR for SST which implies – extrapolating from a handful of sites – that using SST DTR as an estimate of MAT DTR would lead to a systematic underestimate of the MAT DTR. However, the housings used for the sensors at these sites are the same as those used at other sites which do exhibit a clear daytime heating bias, which leaves open the possibility that a smaller heating bias remains in the data from these moorings. It is expected that MAT will be inherently more variable than SST because the air temperature can change more rapidly by advection.

3.2.5 Validation

Ship data for the period 1995-2005 were used to test the relationships. Once again, coincident MAT and SST measurements from the same ship were used. The MAT was predicted using the SST and the offset described above. Two main diagnostics were calculated.

1. The mean and median bias averaged on to a regular grid (to minimise geographical sampling bias that would occur if using individual marine reports).
2. The standard deviation of the normalised "error" equal to the difference between the predicted MAT and the observed MAT divided by the estimated uncertainty. The estimated uncertainty combines the prediction uncertainty with the measurement error uncertainty in the SST measurement (~1°C) and the uncertainty in the MAT measurements (~1°C).

The results of the validation are shown in Figure 3.2.4. Both diagnostics were calculated for every month from 1995 to 2005 (there are 11 orange and blue dots for each month in the figure). The average mean bias is around -0.15°C with a range of +0.2°C to -0.3°C. The mean would, ideally, be zero, but residual uncertainty in the in situ data used to calculate and validate the relationship are expected to be of order 0.1°C and month to month variability can be large even averaged across the globe because the in situ coverage is relatively sparse.
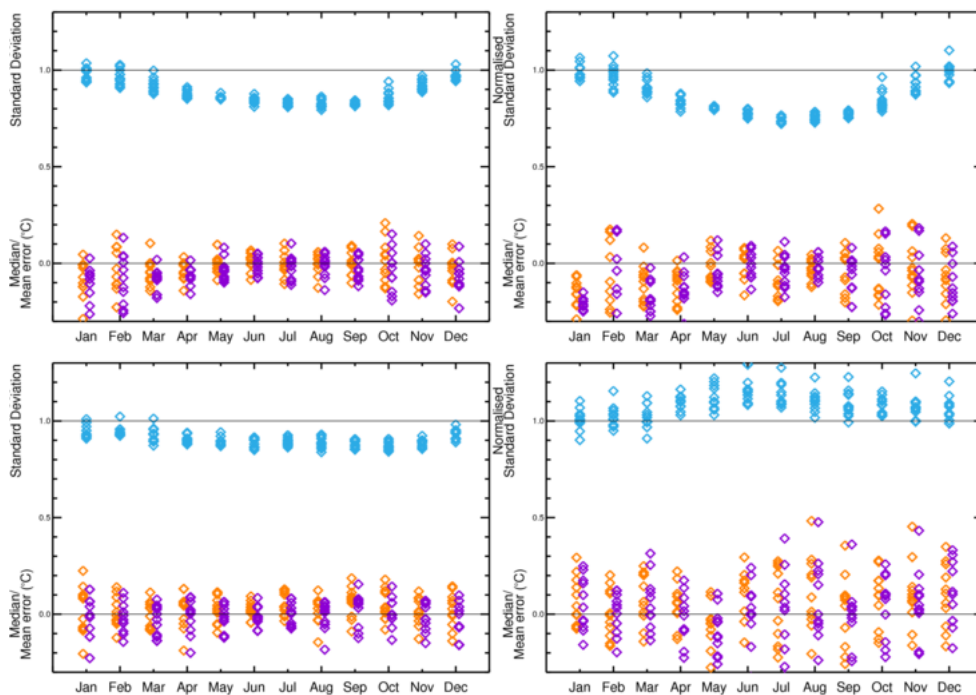
Figure 3.2.4: Testing the relationships using ship data. (top left) global validation results. The blue points are the standard deviations normalised by the estimated uncertainty. The purple and orange points are the median and mean biases respectively. There is one point for each year and month from 1995 to 2005. (top right) Same but for the Northern Hemisphere, (bottom left) tropics, (bottom right) Southern Hemisphere.

The standard deviation of the normalised errors should be around 1 if the uncertainties are correctly specified. An over estimate of the uncertainties would lead to a standard deviation less than 1 and an underestimate of the uncertainty to a standard deviation greater than 1. The estimated uncertainties are therefore slightly overestimated during summer months in the northern

hemisphere, but with a tendency to under-estimation in the Southern Hemisphere. By this measure, errors in the uncertainties are less than 30% on average. The uncertainty is dominated by the locally correlated climatological variability, which is large relative to other components. There is an indication that this uncertainty is underestimated in the southern hemisphere extra tropics, but reasonable elsewhere.

3.2.6 Application to satellite data

SST retrievals from the AATSR instrument for 2002-2012 were used to estimate MAT using the above method. The adjustments were bilinearly interpolated to the target 0.25° grid. We estimate Tmean and ten separate uncertainty fields (see Equation 3.2.3):

1. Uncertainty from random uncorrelated errors from the satellite retrievals.
2. Uncertainty from locally-systematic errors in the satellite SST retrievals which are correlated on approximately synoptic scales i.e. 100km and 1 day.
3. Uncertainty from global systematic errors in the satellite SST retrievals, which are completely correlated in space and time.
4. Uncertainty from locally-systematic errors arising from estimating MAT from SST, which are correlated with a length scale of approximately 1000km.
5. Uncertainty from global systematic errors in the in situ data used to construct the climatology of AST difference. These are correlated in time and strongly correlated in space.
6. -10. Uncertainty in the offset and Fourier components used to smooth the climatology. These are correlated in time (as modulated by the Fourier components), but uncorrelated in space.

Coverage is typically lower than that of the satellite data because relationships between SST and MAT are not specified everywhere. Uncertainties are dominated by the locally correlated error component (number 4 in the list and shown in Figure 3.2.1 for two example days).

3.2.7 Conclusion and outlook

Marine air temperatures (MAT) were estimated from sea-surface temperature (SST) measured in situ and retrieved from satellite measurements. MAT was calculated by adding a spatially and seasonally-varying offset to the SST. The offset was modelled using differences between bias-adjusted SST and MAT measurements made by ships between 1963 and 2000. An uncertainty in the estimated MAT was estimated by analysing the variance of daily deviations from the average offset.

The average offsets were used to predict MAT from SST for ships between 1995 and 2005. The mean bias was around ± 0.2-0.3°C, and the normalised standard deviation of errors was within 30% of the correct value, which indicates that the uncertainties are approximately correct, albeit rather large in some areas of high variability in air-sea temperature difference.

Application to satellite SST retrievals and comparison for the satellite-derived air temperatures to in situ data not used in the estimation of the relationships shows a small warm bias of around 0.1-0.2°C. Part of this is probably due to the fact that the in situ measurements are night time only whereas the SSTs are representative of temperatures at 10.30 am local time which is close the diurnal average. However, it is also consistent with the estimated systematic errors in the satellite and in situ records, which each amount to around 0.1°C.

Future improvements in the bias adjustments applied to in situ data – both SST and MAT – should help to reduce the residual biases in the relationships. There is likely some scope for reducing the size of the estimated uncertainties in high-variability regions such as the Gulf Stream using circulation indices. However, these regions are also well observed by in situ instruments. In other areas, the estimated uncertainties can be much lower – approaching a few tenths of a degree in places – which, even combined with the uncertainty in the satellite-retrieved SSTs is smaller than the typical error in an in situ air temperature measurement. This means that using accurate satellite retrievals of SST it is possible to get a more accurate estimate of the air temperature than can be obtained from a single in situ measurement. This means that the satellite derived MAT and the in situ MAT are likely to be complementary to one another in EUSTACE with satellites providing reliable information in open ocean areas that are only sparsely observed by ships. In areas where the satellite-derived MATs are less reliable owing to high variability of the air-sea temperature difference, such as the Kuroshio and Gulf Stream, there are plentiful in situ observations.

Solar heating biases prevent an accurate estimate of the diurnal temperature range of air temperature, which can likely be remedied only by making more carefully-sited measurements in future.

3.2.8 References

Anderson, Steven P., and Mark F. Baumgartner. "Radiative Heating Errors in Naturally Ventilated Air Temperature Measurements Made from Buoys*." Journal of Atmospheric and Oceanic Technology 15.1 (1998): 157-173.

Atkinson, C. P., N. A. Rayner, J. J. Kennedy, and S. A. Good (2014), An integrated database of ocean temperature and salinity observations, J. Geophys. Res. Oceans, 119, 7139–7163, doi:10.1002/2014JC010053.

Kennedy J.J., Brohan P. and S.F.B.Tett, 2007: A global climatology of the diurnal variations in sea-surface temperature and implications for MSU temperature trends. Geophys. Res. Lett., Vol. 34, No. 5, L05712 doi:10.1029/2006GL028920

Kennedy J.J., Rayner, N.A., Smith, R.O., Saunby, M. and Parker, D.E. (2011c). Reassessing biases and other uncertainties in sea-surface temperature observations since 1850 part 2: biases and homogenisation. J. Geophys. Res., 116, D14104, doi:10.1029/2010JD015220

Kennedy J., Atkinson C., Willett K. (in prep) Quality control of marine observations for climate.

Kent, E.C.; Berry, D.I.. 2008 Assessment of the Marine Observing System (ASMOS): Final Report. Southampton, UK, National Oceanography Centre Southampton, 55pp. (National Oceanography Centre Southampton Research and Consultancy Report 32)

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

Merchant, C. J., Embury, O., Roberts-Jones, J., Fiedler, E., Bulgin, C. E., Corlett, G. K., Good, S., McLaren, A., Rayner, N., Morak-Bozzo, S. and Donlon, C. (2014), Sea surface temperature

datasets for climate applications from Phase 1 of the European Space Agency Climate Change Initiative (SST CCI). Geoscience Data Journal. doi: 10.1002/gdj3.20

Marine Surface Temperature: Observed variations and data requirements, D.E. Parker, C.K. Folland and M. Jackson, Climatic Change 31: 559-600, 1995

Rayner, N.A., P.Brohan, D.E.Parker, C.K.Folland, J.J.Kennedy, M.Vanicek, T.Ansell and S.F.B.Tett 2006: Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: the HadSST2 data set. Journal of Climate. 19(3) pp. 446-469

Woodruff, S.D., S.J. Worley, S.J. Lubker, Z. Ji, J.E. Freeman, D.I. Berry, P. Brohan, E.C. Kent, R.W. Reynolds, S.R. Smith, and C. Wilkinson, 2011: ICOADS Release 2.5: Extensions and enhancements to the surface marine meteorological archive. Int. J. Climatol. (CLIMAR-III Special Issue), 31, 951-967 (doi:10.1002/joc.2103).

3.3 Relationships over land

3.3.1 Introduction to land relationships

Efforts to estimate land surface air temperature (LSAT) from satellite observations have been widely reported in the literature, with a notable increase in the number of publications on this subject in recent years (Benali et al., 2012; Chen et al., 2014; Good, 2015; Janatian et al., 2016; Kilibarda et al., 2014; Oyler et al., 2015; Parmentier et al., 2015). This is largely due to: an increasing need to deliver LSAT data sets at higher spatial resolution; improvements in the quality and availability of satellite land surface temperature (LST) data (Trigo et al., 2008; Freitas et al., 2010, Freitas et al., 2013; Wan, 2013; Wan 2016); and an improved understanding of the relationship between LST and LSAT (Gallo et al., 2011; Good, 2016; Mildrexler et al., 2011). However, other than Atmospheric Infrared Sounder (AIRS) atmospheric temperature profiles, which include a 12-hourly instantaneous estimate of the near-surface air temperature (AIRS Science Team/Joao Texeira, 2013), there are currently no multi-year, global satellite LSAT data sets in the public domain. Since the AIRS data do not represent the daily extremes, Tmin and Tmax, and given the 50-km spatial resolution of these data - approximately a quarter of the EUSTACE target spatial resolution - a new satellite-derived LSAT product has been developed for EUSTACE.

Nearly all previous studies concerned with estimating satellite-derived LSAT have adopted a statistical empirical approach to predicting LSAT from satellite infrared LST and other variables, such as vegetation and elevation (Benali et al., 2012; Chen et al., 2014; Good, 2015; Janatian et al., 2016; Oyler et al., 2015). These studies have demonstrated that it is possible to achieve root-mean squared differences (RSMD) between satellite-derived and station-measured LSAT observations of around 2-5 °C using this method, which is reasonable given the 1-3 °C accuracy of the current satellite LST data (Dash et al., 2002; Freitas et al., 2010, Freitas et al., 2013; Li et al, 2013; Trigo et al., 2008) and inherent discrepancies in comparing point station and areal-averaged satellite data, which may amount to ~2 °C (K. Veal, University of Leicester, Personal Communication). The approach adopted in EUSTACE has been to build on previous research in this field and develop a simple empirical statistical model that can be applied on a multi-decadal, global scale, without excessive computational expense. To our knowledge, such a task has only

previously been attempted by Kilibarda et al. (2014), who estimated global daily LSAT for one year using satellite LST data from MODIS and spatio-temporal regression-Kriging. The development of the EUSTACE satellite LSAT product is described in the following sections. Section 3.3.2 introduces the data sets used in the study. The variability of the LST-LSAT relationship is explored briefly in section 3.3.3, while the LSAT statistical model is presented in sections 3.3.4 and 3.3.5. The model is evaluated in section 3.3.6, and concluding remarks are outlined in section 3.3.7. The feasibility of using a physical model as an alternative approach to estimating LSAT using satellite data is explored in a later section (section 3.6).

3.3.2 Data

Table 3.3.1 provides an overview of the data sets used in the study. Other than the data from the Atmospheric Radiation Measurement (ARM) sites, all data are used in the prediction of LSAT using the empirical statistical model.

Table 3.3.1: *Data sources used in this study. The uncertainty information provided for each data set is also shown; 'full breakdown' indicates that the information is partitioned into e uncertainties with different correlation properties.*

Variable	Source	Native spatial resolution	Native temporal resolution	Uncertainty information
LSTday and LSTngt	GlobTemperature MODIS Aqua (Wan, 2013; Wan 2016)	~ 1 km	Daily	Full breakdown, as derived in earlier EUSTACE work (see Merchant et al., 2015 and 2017)
FVC Fraction of vegetation	Copernicus/Geoland-2 global land service FCOVER (*Baret et al.*, 2013; *Camacho et al.* 2013)	1/112°	10-day	Total only
DEM Digital Elevation Model	Shuttle Radar Topography Mission v1 (*Farr & Kobrick,* 2000; *Rodriguez et al.,* 2005)	1 km (down-sampled from 3 arc second)	N/A	None
Snow fraction	NASA MODIS Aqua (Hall et al., 2006)	0.05°	Daily	Quality and confidence (0-100) indices only
Station Tmin & Tmax	GHCN-D v3.22 (Menne et al., 2012)	N/A	Daily	Quality flags only
Station Tmin & Tmax	Eustace Station Version R000771 (EUSTACE Deliverable D1.7)	N/A	Daily	Quality flags only
Land Cover type	LC CCI (v1.6) 2005 epoch (Bontemps et al., 2012; Poulter et al., 2015)	300m	N/A	Algorithmic confidence level (0-100)
Station LST, LSAT, wind and cloud	ARM (Mather and Voyles, 2013; Stokes and Schwarz, 1994)	N/A	1-minute	Quality flags only

3.3.3 Variability of the LST vs. LSAT relationship

LST and LSAT are often well coupled, and the two temperatures can be close (within 1°C) under certain conditions, for instance in the presence of densely vegetated canopies such as forests, or when insolation is low or absent, for example under cloud, at night or at high latitudes (Good, 2016; Hachem et al., 2012; Mildrexler et al., 2011; Sohrabinia et al., 2014; Urban et al., 2013; Vancutsem et al., 2010). However, LST and LSAT frequently differ by several degrees or more, with the largest differences occurring close to the peak of the diurnal cycle under cloud-free skies (Duda and Minnis, 2000; Edwards et al., 2011; Gallo et al., 2011; Good, 2015; Jin et al., 1997; Jin & Dickinson, 2010). Maximum LST has been found to be as much as 20 degrees above the corresponding maximum LSAT in extreme cases (Mildrexler et al., 2011).

Figure 3.3.1 (reproduced from Good, 2016) shows examples of the diurnal temperature evolution over a 24 hour period from six deployments of the Atmospheric Radiation Measurement (ARM) programme (Stokes and Schwartz, 1994; Mather and Voyles, 2013). Figure 3.3.1a, showing data from the Niamey site on a virtually cloud-free day, is an example of what is often considered to be the classic diurnal cycle of LST and LSAT: the two temperatures are very close at night, while LST is warmer than LSAT during the day with maximum LST (LSTmax) occurring before Tmax. This relative temperature behaviour is well known and is similar to examples presented in other studies (Edwards et al., 2011; Jin et al., 1997; Jin and Mullins, 2014). However, most diurnal cycles do not exhibit this "classic" behaviour owing to the influence of cloud, wind speed and changes in local meteorology. The effects of these factors on the LST-LSAT relationship are clear in the other examples shown in Figure 3.3.1. In general, the presence of cloud and/or increased wind during the day reduces LST and aligns it more closely with LSAT.

Figure 3.3.2 shows a time series of daily MODIS LST minus GHCN-D daily maximum (Tmax) and minimum (Tmin) air temperature over Europe during 2010. The local observation time for the MODIS LST data is nominally 1:30 am/pm. Differences are shown for LSTday-Tmax and LSTngt-Tmin, partitioned into surfaces where the vegetation fraction is ≤25% (sparse), 25-75% (moderately vegetated) and ≥75% (fully vegetated). It is notable that the LSTngt-Tmin difference is smaller in magnitude and the relationship more stable in time than the LSTday-Tmax time series. LSTngt tends to be slightly warmer than Tmin. This is expected because LSTngt is observed at ~ 01:30 local solar time and Tmin, usually occurs around sunrise (Good, 2016). The closeness and stability of the LSTday-Tmax relationship is greatest during winter months, whereas the largest differences and greatest instability occur during the summer. Vegetation clearly has a strong influence on the LST-LSAT relationship, with smaller-magnitude and more stable differences in the presence of vegetated surfaces. In the presence of fully-vegetated surfaces, there is little difference between the median LSTngt and Tmin. These observations indicate that both vegetation and time of year (or solar zenith angle) have a strong influence on the LST-LST relationship, particularly between LSTday and Tmax.

Over snow-covered surfaces, the LST-LSAT difference tends to be closer to zero. Figure 3.3.3 (reproduced from Good, 2016) shows the variation in LST-LSAT with satellite-estimated snow fraction, also from MODIS Aqua (Hall et al., 2006). For LSTmin-Tmin there is no clear relationship with snow cover. However, for LSTmax-Tmax, the difference can be large without snow cover, but tends towards zero even with small snow cover fraction.

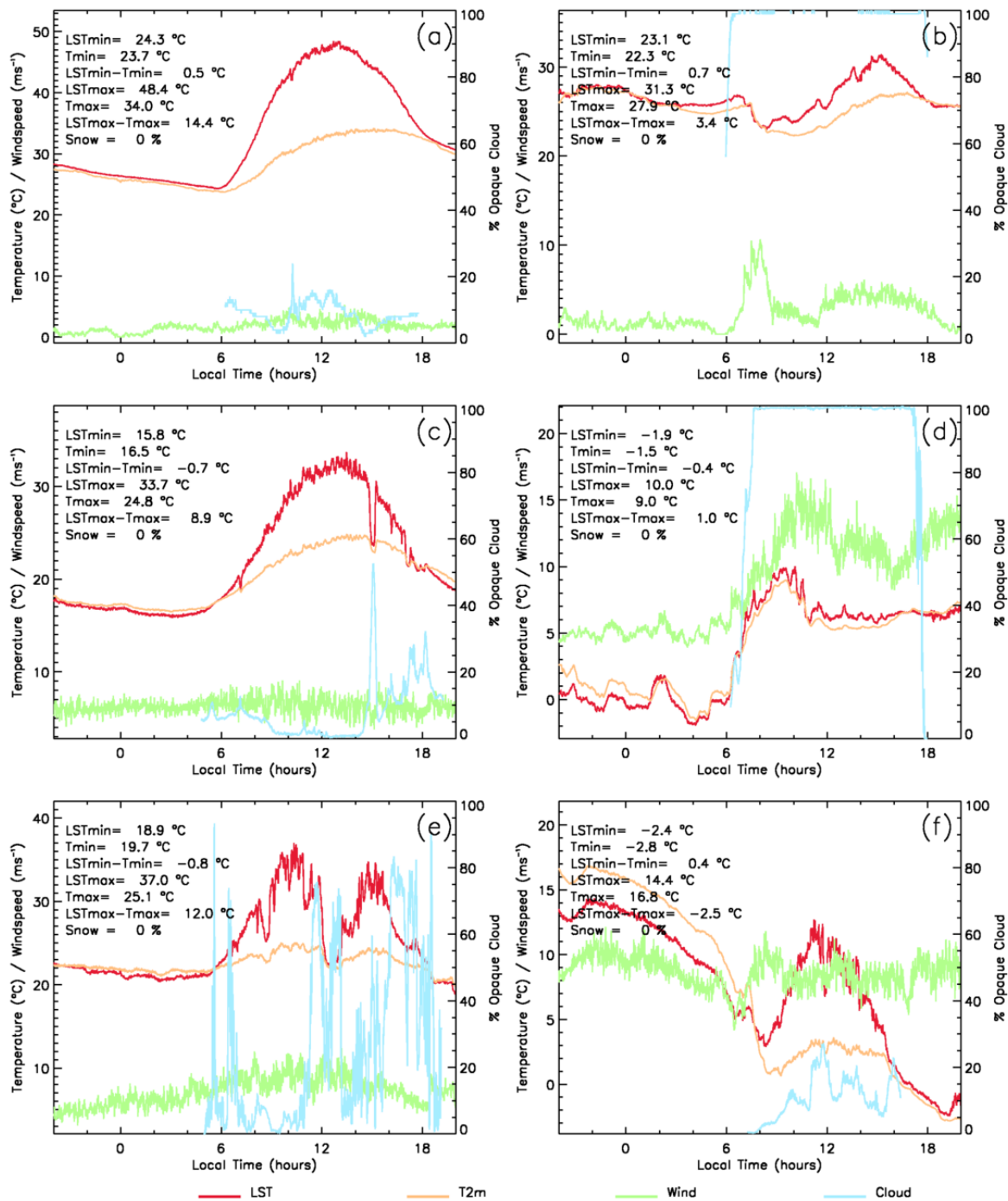**Figure 3.3.1:** *Examples of diurnal evolutions of LST (°C, red) and LSAT (°C, orange), wind speed (m/s, green) and % opaque cloud (blue) measured in situ at the ARM sites at (a) Niamey (NIM) on 2 September 2006, (b) Niamey on 9 June 2006, (c) Cape Cod, Massachusetts (PVC) on 21 June 2013, (d) Oklahoma (SGP) on 26 March 2014, (e) Graciosa Island, Azores (ENA) on 25 June 2014 and (f) Oklahoma (SGP) on 30 November 2014.*

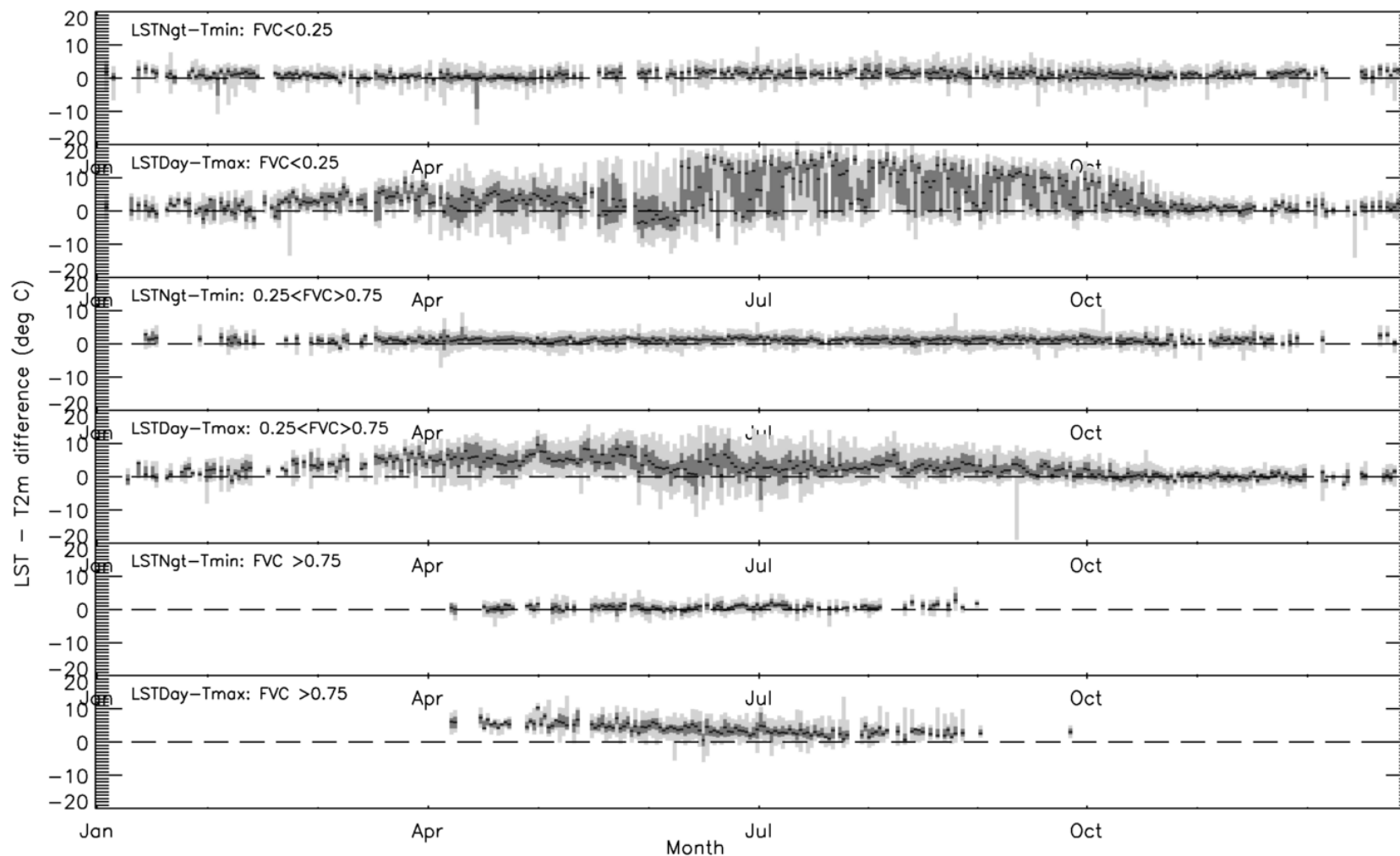**Figure 3.3.2:** *Time series of LST minus LSAT (T2m) distributions (°C) for differing fraction of vegetation (FVC) within the satellite pixel during 2011 over Europe (34 to 75°N, -12 to 40 °E). The 50th (black), 25-75% centile (dark grey) and 5-95th centile (light grey) ranges are shown on the plot.*

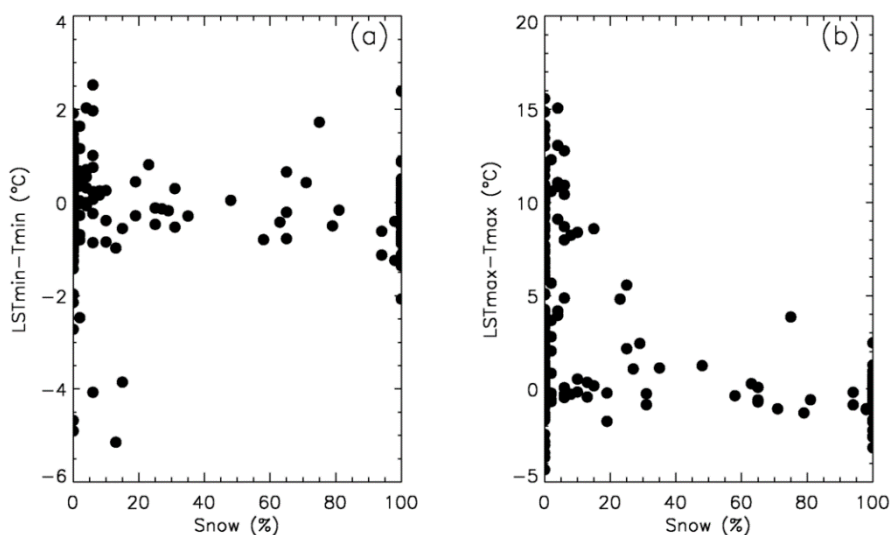EUSTACE (640171) Deliverable 1.5   Page 25

Figure 3.3.3: *Relationship between LST-LSAT (°C) and snow cover (% snow in satellite pixel) for (a) minimum and (b) maximum temperatures for Olitok Point (OLI) during 2014.*

3.3.4 Relationship model

The method for estimating LSAT from satellite data is based on the approach detailed in Good (2015) and references therein, which was developed for data from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) over Europe. The model is essentially a multiple linear regression used to predict daily *Tmin* and *Tmax* from daily satellite LST and other predictors, such as vegetation.

The generalised equations for predicting *Tmin* and *Tmax* are:

$$Tmax = a_0 + a_1.LST_{day} + a_2.LST_{ngt} + a_3.FVC + a_4.SZA_{noon} + a_5.Snow + \varepsilon_{Tmax.} \qquad \text{(Eq. 3.3.1)}$$

$$Tmin = \beta_0 + \beta_1.LST_{day} + \beta_2.LST_{ngt} + \beta_3.FVC + \beta_4.SZA_{noon} + \beta_5.Snow + \varepsilon_{Tmin} \qquad \text{(Eq. 3.3.2)}$$

Where a and β are the multiple linear regression coefficients. LST_{day} and LST_{ngt} are the retrieved daytime and night-time satellite LSTs (in °C), respectively. *FVC* is the fractional vegetation cover (0-1), *Snow* is the % snow cover and SZA_{noon} is the solar zenith angle at local solar noon (in degrees). ε is the error on the model. All quantities correspond to the mean value for each grid cell at the specified spatial resolution. The

regression coefficients are given in Table 3.3.2, together with the valid range of data for each predictor. Model error terms are discussed further in Section 3.3.5.

A global approach is adopted, whereby one set of coefficients is used for every spatial point at all times. Three model variants are used, which require different sets of predictors. The primary model for both temperature predictions, $Tmin_1$ and $Tmax_1$, requires both LST_{day} and LST_{ngt} in each case. The secondary and tertiary models - $Tmin_2$, $Tmax_2$, $Tmin_3$ and $Tmax_3$ – require only one LST to be present, thereby enabling greater spatial coverage of the LSAT predictions because LST_{day} and/or LST_{ngt} may be unavailable on a particular day, e.g. due to cloud or failed observations. The primary models are used wherever possible as these are the most accurate. The secondary or tertiary models are used for other locations; it should be noted that these are mutually exclusive, so are not applied in any preferential order. The global fields of *Tmin* and *Tmax* are therefore a 'best guess' field and may include data from all three models.

Table 3.3.2: *Model regression coefficients and the valid data range for each predictor, determined physically (FVC, Snow, SZA) or from the input data (LST). 'StDev' gives the standard deviation of the model residuals (i.e. predicted satellite-derived LSAT minus observed station LSAT).*

Model	Offset (°C) α_0 / β_0	LSTday α_1 / β_1	LSTngt α_2 / β_2	FVC α_3 / β_3	SZAnoon α_5 / β_5	Snow α_6 / β_6	StDev (°C)
Tmin1	-1.513	0.032	0.835	0.765	0.000	0.000	2.84
Tmax1	7.092	0.388	0.432	1.516	0.000	-0.011	3.02
Tmin2	0.184	0.000	0.850	0.595	-0.021	0.000	2.84
Tmax2	5.042	0.594	0.000	2.956	0.000	-0.022	3.65
Tmin3	-5.734	0.436	0.000	3.601	0.000	0.000	4.88
Tmax3	21.260	0.000	0.723	0.00000	-0.130	-0.055	3.88
Range$_{min}$	-	-80.0 °C	-80.0°C	0.0	0.0°	0	-
Range$_{max}$	-	65.0 °C	40.0°C	1.0	90.0°	100	-

3.3.4.1 *Choice of predictors*

The choice of predictors is based on current knowledge of the parameters that influence the relationship between LST and LSAT (section 3.3.3; Good, 2016; Hachem et al., 2010; Jin et al., 1997; Mildrexler et al., 2011; Urban et al., 2013), and other studies published in the literature that have estimated LSAT using satellite data (Benali et al., 2012; Chen et al., 2014; Good, 2015; Kilibarda et al., 2014; Oyler et al., 2015; Parmentier et al., 2015).

The correlation between all predictor variables tested in the study is shown in Table 3.3.3. Both *Tmin* and *Tmax* are well correlated with both *LSTday* and *LSTngt* ($r \geq |0.72|$). Unsurprisingly, the temperature variables are also quite well correlated with SZA_{noon} ($r \geq |0.62|$). The correlation between both FVC and DEM and temperature is weak ($r < |0.22|$), while snow is moderately well correlated with temperature ($r \sim 0.33$).

Table 3.3.3: *Correlation between variables tested in model training. All correlation coefficients have p values <0.01. The number of data pairs in each case is 98,773.*

Model	Tmin	Tmax	LSTday	LSTngt	FVC	DEM	SZAnoon	Snow
Tmin	1.00	0.82	0.72	0.92	0.17	-0.22	-0.62	-0.30
Tmax	0.82	1.00	0.89	0.86	0.12	-0.10	-0.70	-0.35
LSTday	0.72	0.89	1.00	0.77	-0.01	0.03	-0.78	-0.35
LSTngt	0.92	0.86	0.77	1.00	0.16	-0.21	-0.64	-0.31
FVC	0.17	0.12	-0.01	0.16	1.00	-0.13	-0.13	-0.05
DEM	-0.22	-0.10	0.03	-0.21	-0.13	1.00	-0.13	0.06
SZAnoon	-0.62	-0.70	-0.78	-0.64	-0.13	-0.13	1.00	0.26
Snow	-0.30	-0.35	-0.35	-0.31	-0.05	0.06	0.26	1.00

The impact of each predictor in the model has been assessed by analysing the model residuals (predicted minus actual) through an incremental addition of parameters. The impact on model residuals partitioned by surface type, e.g. FVC ≥0.7, 0.7>FVC>0.3, and FVC ≤0.3 is also examined. This is an important step in assessing the impact of predictors since the stations used for training may not accurately represent the proportions of each surface regime in the real world so the impact may not be apparent in the overall residual statistics. For example, the impact of including FVC in the model to predict *Tmax* is negligible when considering the full set of residuals, but when considering different FVC categories the impact is notable. In this case, the addition of *FVC* for those space-time locations with high *FVC* (≥0.7), a reduction in the mean absolute difference (MAD) from 1.1 °C to 0.5 °C is observed, while for those with mid (0.7>FVC>0.3) and low *FVC* (≤0.3) the reduction in MAD is from 0.4 °C to 0.3 °C, and 0.7 °C to 0.3 °C, respectively.

Using both LST_{day} and LST_{ngt} as predictors in the *Tmax* model reduces the standard deviation of the residuals by >0.5 °C, compared to using just LST_{day}. This is likely to be because LST_{day}-LST_{ngt} difference is sensitive to the thermal properties of the land surface, and synoptic conditions, and thus the LSAT diurnal temperature range, therefore improving the ability of the model to predict Tmax. However, as LST_{ngt} and *Tmin* are already well coupled, the impact of using LST_{day} in the *Tmin* model is very small, resulting in a reduction

in standard deviation of <0.1 °C. However, magnitude of the LST_{day} term can be a few tenths of a °C. Additionally, LST_{ngt} is more susceptible to cloud contamination than LST_{day}, where visible channel data are available, so including LST_{day} acts as a stabilising factor on the estimation of LSAT.

3.3.4.2 Model training

Model training is performed using satellite data aggregated to 0.05 °lat/long, which is at a finer spatial resolution that the EUSTACE target grid of 0.25 °lat/long. We chose to train the model at the higher resolution because the area-averaged LST vs point LSAT relationship peaks at around 5 km spatial resolution [Sohrabinia et al., 2014], and degrades with decreasing spatial resolution. Moreover, more satellite-station matchups are obtained using a higher resolution, providing a larger data pool on which to perform the model training.

The model is trained using station observations from EUSTACE Station data set. This global data set includes around 20,000 stations that observe *Tmin* and/or *Tmax* over the MODIS era (from the year 2000 onwards). A sub-set of 2206 stations is used here to generate the coefficients for estimating satellite-derived LSAT, leaving the remaining stations for potential use within the EUSTACE global surface air temperature analysis or for validation. Stations are preferentially chosen for model training if: the station land cover type matches the dominant land cover type of the grid cell containing the station; it has the greatest number of years of data available between 2000 and 2016 from available station choices; the elevation of the station matches the mean elevation of the grid cell. To minimise well-observed regions dominating the model training, some station thinning is also performed using these same criteria.

The LST data are retrieved from observations made at infrared wavelengths and are therefore only available when conditions between the sensor and the surface are cloud free. The GlobTemperature LST products include quality information that indicates when retrievals are cloud-contaminated. However, cloud detection is a non-trivial problem and cloud is sometimes 'missed', which can lead to large errors in the retrieved LSTs. This error is nearly always a cold bias, such that the cloud-contaminated LST retrievals are cold relative to the true LST by up to several °C. Cloud contamination is often most prevalent at night, where the absence of data from visible channels results in more challenging conditions for detecting cloud successfully. The cloud mask provided with the GlobTemperature LST data is binary rather than probabilistic, indicating either 'cloud' or 'no cloud'. The uncertainties estimated for the retrieved LSTs do not include any allowance for possible cloud effects.

In order to minimise the impact of cloud-contaminated LSTs on the regression coefficients, additional screening is performed by sub-sampling the data set to include only the day with the maximum LST over successive non-overlapping 10-day windows. Data (both satellite and *in situ*) corresponding to the other nine days within the window are excluded from the analysis. As cloud causes a cold-biased LST, this 10-day maximum LST is the least likely observation to be contaminated by cloud. Since the data record is long (> 10 years) and the coverage is global, this will not result in a warm biased training data set because colder conditions are still captured at higher latitudes and during cooler seasons at all latitudes. This process increases the correlation coefficients between predictors and the multiple linear regression coefficient, indicating this has a positive effect on the analysis. Grid cells (0.05 °lat/long) containing fewer than 50% cloud-free 1km LST observations or a sampling uncertainty greater than 5 °C are also excluded from the pool of data used for regression in order to minimise cloud contamination on the data used in the relationship building. The sampling uncertainty represents the uncertainty within the grid cell arising from incomplete sampling of LST retrievals within the grid cell; it is a function of both the cell variance and the number of available observations. Sampling uncertainty is often found to be inflated where observational coverage is poor, and/or where cloud contamination is present in part of the cell, owing to the presence of cold-biased LSTs.

3.3.4.3 Model implementation

While the model coefficients are derived using data at 0.05°, the model is implemented at 0.25° to match the EUSTACE target grid. The implementation of the model is straightforward. For each 0.25° cell, the predictor variables are aggregated and uncertainties are propagated (where possible). The coefficients in Table 3.3.2 are then applied to each predictor variable to give an estimate of Tmin and Tmax for that cell, according to Equations 3.3.1 and 3.3.2. The primary model – Model 1 – is implemented wherever possible, followed by Model 2 and Model 3 (which are mutually exclusive). LSAT is not estimated for grid cells where the percentage of cloud-free 1 km pixel LST observations within that grid cell is less than 20%, or where the sampling uncertainty is greater than 3 °C. This is to reduce the amount of residual cloud contamination in the satellite LSAT estimates, as discussed above. The choice of these thresholds is a compromise between successfully rejecting poor-quality data while still preserving good data coverage. LSAT is also not estimated for cells classified as land or sea ice; near-surface air temperatures for these cells are provided in the ice air temperature products described in Section 3.4. Figures 3.3.4 and 3.3.5 show an example of the fields for a single calendar day.
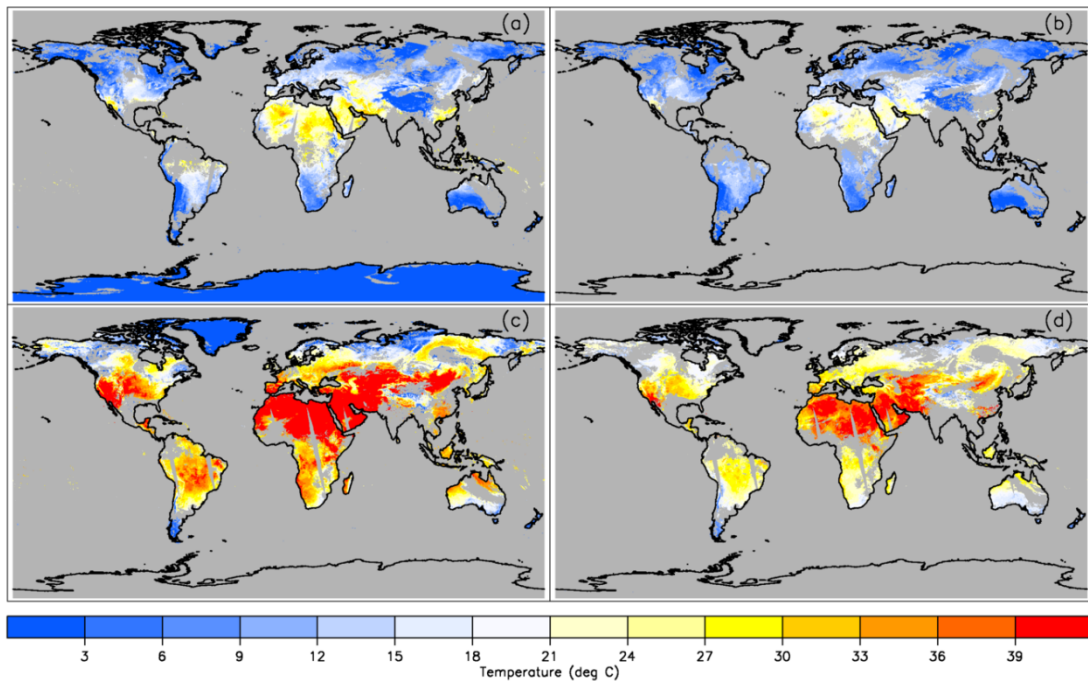
Figure 3.3.4: Maps showing (a) satellite-retrieved LST$_{ngt}$,(°C) (b) satellite-predicted Tmin (°C), (c) satellite-retrieved LST$_{day}$, (°C) and (d) satellite-predicted Tmax (°C) on 1 July 2010. Grey areas correspond to missing data, e.g. due to cloud, ice or ocean.
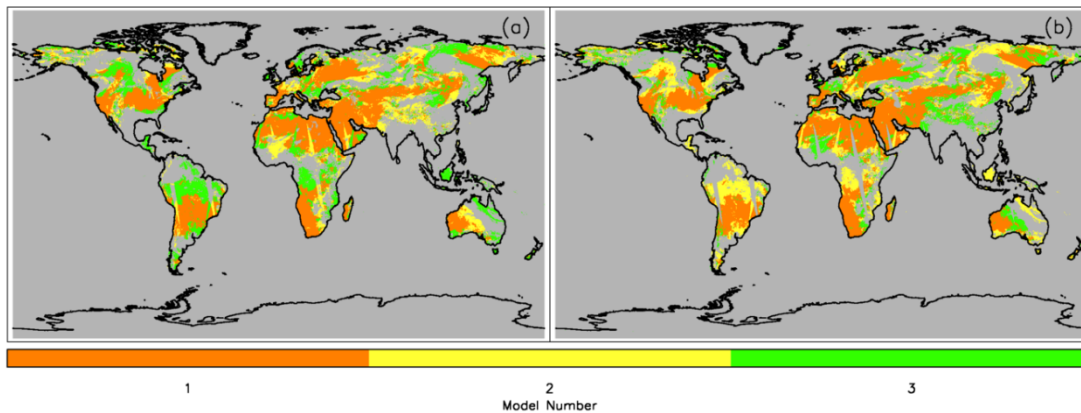
Figure 3.3.5: *Maps showing the model number used for satellite-predicted (a) Tmin and (b) Tmax on 1 July 2010. Grey areas correspond to missing data, e.g. cloud, ice, or ocean.*

The temporal coverage of the satellite-derived LSAT product exceeds 50 % for most of the globe (not shown). As expected, temporal coverage is highest in the clear-sky arid regions, for example North Africa, and lowest in the tropics, where cloud is most prevalent.

3.3.5 Modelling uncertainties

Uncertainties are estimated for each predictor variable based on the uncertainty information available for each native data set used (Table 3.3.1). For the LST data – which is the only predictor to have complete characterisation of uncertainties - each uncertainty component (random, locally correlated, systematic) is propagated through the gridding process according to EUSTACE Deliverable 1.2 (Merchant et al., 2015). For the *FVC* data set, the mean total uncertainty within a particular grid cell is used as an estimate of the locally correlated uncertainty for that grid cell, while the difference between the maximum and the mean uncertainties within each cell is used as an estimate of the random uncertainty. The quoted precision of the FVC estimates, 0.004, is assumed to be the systematic uncertainty on a pixel-by-pixel basis. At present, no uncertainty information has been estimated for the *snow* estimates. As the dominant contribution to the satellite-derived LSAT estimates is from the LST data, omitting the uncertainty contribution from the snow data should have only a very small impact.

Uncertainty estimates for the satellite-derived LSAT data are estimated so that for a particular grid cell at a particular point in time, the random uncertainty component is given by:

$$Tmax_{random} = (\alpha_1^2 . LST_{day\_random}^2 + \alpha_2^2 . LST_{ngt\_random}^2 + \alpha_3^2 . FVC_{random}^2)^{\frac{1}{2}} \qquad \text{(Eq. 3.3.3)}$$

$$Tmin_{random} = (\beta_1^2 . LST_{day\_random}^2 + \beta_2^2 . LST_{ngt\_random}^2 + \beta_3^2 . FVC_{random}^2)^{\frac{1}{2}} \qquad \text{(Eq. 3.3.4)}$$

Where LST_{day_random}, LST_{day_random} and FVC_{random} are the random uncertainty components from the input data sets for the same grid cell at the same point in time for LST_{day}, LST_{night} and FVC, respectively. The coefficients α and β are the model coefficients from Table 3.3.2; so for models that do not use a particular predictor, that term in equations 3.3.3 and 3.3.4 will be zero. Similarly, the locally correlated uncertainty components are given by:

$$Tmax_{atm} = (\alpha_1^2 . LST_{day\_local\_atm}^2 + \alpha_2^2 . LST_{ngt\_loca\_atml}^2 + \sigma_{Tmax}^2)^{\frac{1}{2}} \qquad \text{(Eq. 3.3.5)}$$

$$Tmin_{atm} = (\beta_1^2 . LST_{day\_local\_atm}^2 + \beta_2^2 . LST_{ngt\_local\_atm}^2 + \sigma_{Tmin}^2)^{\frac{1}{2}} \qquad \text{(Eq. 3.3.6)}$$

$$Tmax_{surf} = (\alpha_1^2 . LST_{day\_local\_surf}^2 + \alpha_2^2 . LST_{ngt\_local\_surf}^2 + \alpha_3^2 . FVC_{local}^2)^{\frac{1}{2}} \qquad \text{(Eq. 3.3.7)}$$

$$Tmin_{local} = (\beta_1^2 . LST_{day\_local\_surf}^2 + \beta_2^2 . LST_{ngt\_local\_surf}^2 + \beta_3^2 . FVC_{local}^2)^{\frac{1}{2}} \qquad \text{(Eq. 3.3.8)}$$

Where $LST_{day\_local\_atm}$, $LST_{ngt\_local\_atm}$, $LST_{day\_local\_surf}$, $LST_{ngt\_local\_surf}$ and FVC_{local}, are the locally correlated uncertainty components from the input data sets for the same grid cell at the same point in time for LST_{day}, LST_{night} and FVC, respectively. (Note: The GlobTemperature MODIS LST products provide separate uncertainty estimates due to atmospheric (atm) and surface (surf) effects on the LST retrievals, and it is these separate components that are used here.) The terms σ_{Tmin} and σ_{Tmax} represent the standard deviation of the residuals, i.e. the satellite-predicted Tmin/Tmax minus the station Tmin/Tmax, calculated during the model training process ('StDev' in Table 3.3.2). The grid-cell systematic uncertainties are assumed to be 0.1 °C, which is consistent with the systematic uncertainty estimated for the GlobTemperature LST data.

3.3.6 Validation

Formal evaluation of all EUSTACE products is being performed in WP3. However, evaluation of the LSAT model performance is presented here using ~7000 EUSTACE stations not used in the model training (section 3.3.4.1). The evaluation is performed for satellite-derived LSAT estimates at 0.05 ° latitude/longitude between July 2002 and

December 2016. The satellite-derived LSAT data have been sub-sampled to reduce the effect of cloud, as previously described (section 3.3.4.3). A summary of the global performance is shown in Table 3.3.4. Figures 3.3.6 and 3.3.7 present a graphical illustration of the performance of Model 1.

Table 3.3.4: *Summary evaluation results for each LSAT model using ~7000 stations not used in model training. 'Correlation' indicates the correlation between all station-observed and satellite-predicted LSAT pairs, while 'slope' indicates the slope of a linear regression line fitted to pairs of satellite-derived air temperatures and independent station measurements. The median and root-mean squared differences (RMSD) are also given.*

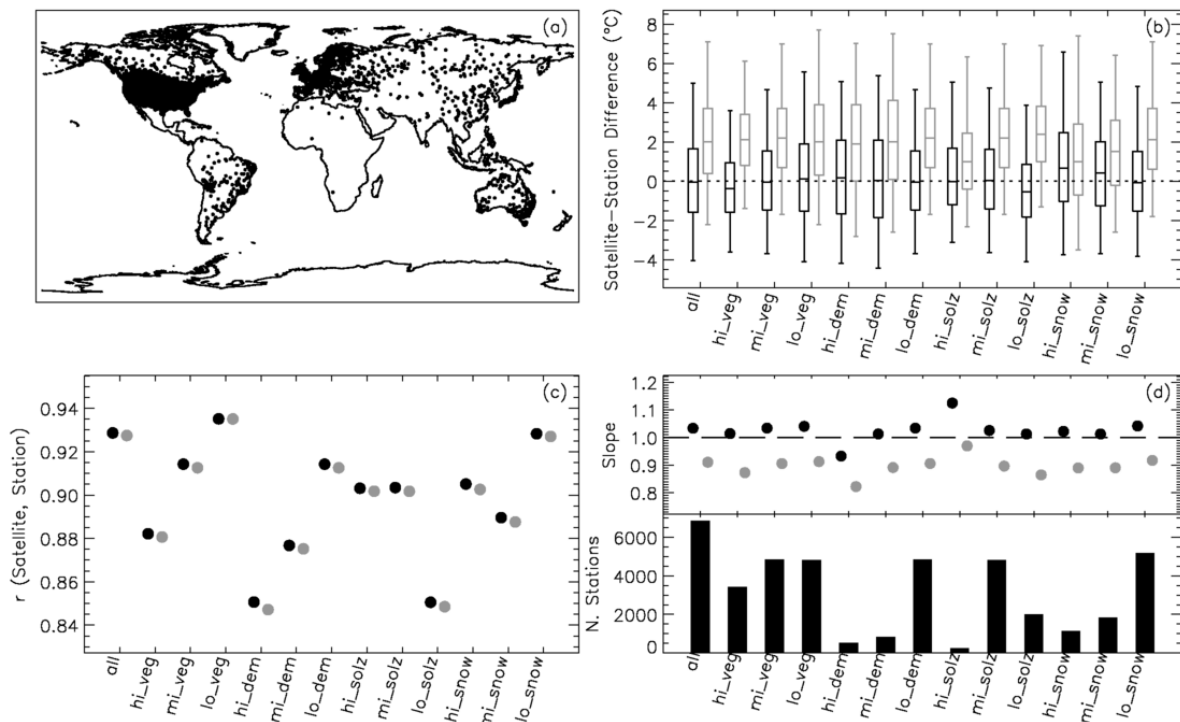| Model	Correlation	Median (°C)	RSMD (°C)	Slope
Tmin1	0.93	-0.04	2.80	1.03
Tmin2	0.93	-0.03	2.80	1.03
Tmin3	0.76	0.10	4.89	1.07
Tmax1	0.93	0.00	3.08	1.00
Tmax2	0.89	-0.07	3.76	1.01
Tmax3	0.88	0.06	3.89	0.99

Figure 3.3.6: *Global Tmin evaluation results (~7000 stations) for Model 1. In panels b-d, the grey bars/filled circles represent results for night time satellite LST minus station Tmin; the black bars/filled circles correspond to satellite-estimated Tmin minus station Tmin. Panel (a) shows validation station locations (filled black circles), (b) shows satellite minus station difference where the whiskers correspond to the 5th/95th centiles and the box shows the 25th/75th centiles with the median shown by the horizontal line within the box, (c) shows the satellite vs. station correlation coefficient and (d) shows the station vs. satellite linear slope (top) and the number of stations in category (bottom). Results are presented for all stations (all), stations with high (FVC≥0.7), mid (0.3<FVC<0.7) and low (FVC≤0.3) FVC, stations with high (DEM>2000), mid (1000≤DEM≤2000) and low (DEM<1000) elevation, stations with high (SZA≥75°), mid (20≤SZA≤75) and low (SZA≤20°) SZA, and stations with high (snow≥75°), mid (20≤snow≤75) and low (snow≤20°) snow %.*
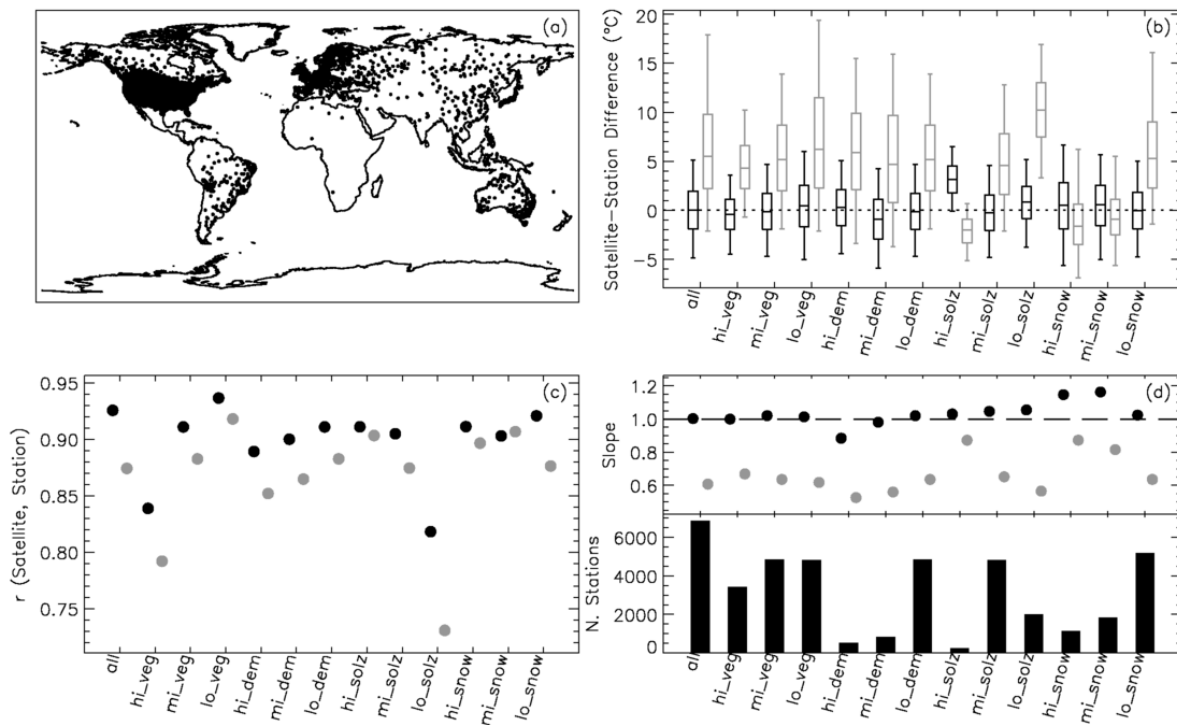
Figure 3.3.7: *As for Figure 3.3.6 but for Model 1 Tmax evaluation. Grey bars/circles denote results using LST_day.*

The summary results presented in Table 3.3.4 are encouraging and in line with the results of previous studies (Benali et al., 2012; Chen et al., 2014; Good, 2015; Janatian et al., 2016; Kilibarda et al., 2014; Oyler et al., 2015; Parmentier et al., 2015). The performance of the primary models (Tmin1 and Tmax1) is generally superior to the secondary and tertiary models, although it is notable that there is no real difference in the performance of Tmin1 (LSTday and LSTngt) and Tmin2 (LSTngt only). However as noted earlier, using Tmin1 may be preferable because this uses LSTday, which is less susceptible to cloud contamination than LSTngt. Examining the results presented in Figures 3.3.6 and 3.3.7 indicates that the LSAT models provide better estimates of LSAT than using LST as a proxy for LSAT. The results also suggest that model performance varies somewhat with surface regime. For example, Tmax estimates for matchups with a high solar zenith angle (e.g. high latitude locations) do not perform well, which is likely to be due to the very low number of stations in this regime. It should be noted that Section 3.4 discusses relationships specifically for ice-covered surfaces and these will be used in EUSTACE products here.

3.3.7 Conclusions and outlook

A computationally inexpensive, simple approach for estimating global LSAT from satellite data has been proposed for the EUSTACE project. The method predicts Tmin and Tmax using linear models that depend on LST_{day}, LST_{ngt}, FVC, and SZA_{noon}. Separate models are proposed for Tmin and Tmax, with three model variants in each case that rely on different combinations of predictors in order to maximise spatial coverage. The method can be applied at locations (and at any spatial resolution) with at least one cloud-free LST observation on a given day. An uncertainty model is proposed, whereby four uncertainty components are provided with each satellite LSAT estimate: random, locally correlated (atmospheric), locally correlated (surface) and systematic. These can be combined in quadrature to provide a total uncertainty.

The accuracy of the LSAT estimates has been independently assessed at around 7,000 stations. The Tmin predictions are found to be more accurate than those of Tmax. The median satellite-station difference ranges between -0.07 and 0.10 °C, with a RMSD of between 2.80 and 4.89 °C, depending on the model used. The primary models – which use both LSTday and LSTngt - achieve a median difference and RMSD of -0.04 and 2.80 °C, respectively for Tmin and 0.00 and 3.08 °C for Tmax.

Spatial coverage of the product could be improved in future by using LST retrievals from more than one sensor, for example MODIS Terra. A new set of model coefficients would need to be generated for each new sensor used, as the relationship between LST and Tmin/Tmax will vary with sensor overpass time.

3.3.8 References

Baret, F., Weiss, M., Lacaze, R., Camacho, F., Makhmara, H., Pacholcyzk, P., & Smets, B. (2013). GEOV1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part1: Principles of development and production. Remote Sensing of Environment, 137, 299-309

Benali A, A. C. Carvalho, J. P. Nunes, N. Carvalhais, A. Santos (2012), Estimating air surface temperature in Portugal using MODIS LST data, Remote Sens. Environ. 124: 108–121, doi: 10.1016/j.rse.2012.04.024.

Bontemps, S., M. Herold, L. Kooistra, A. van Groenestijn, A. Hartley, O. Arino, I. Moreau and P. Defourny (2012), Revisiting land cover observations to address the needs of the climate

modelling community. Biogeosciences Discussions, 8(4), 7713-7740 (doi:10.5194/bg-9-2145-2012).

Camacho, F., Cernicharo, J., Lacaze, R., Baret, F., and Weiss, M. (2013). GEOV1: LAI, FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part 2: Validation and intercomparison with reference products. Remote Sensing of Environment, 137, 310-329

Chen, F., Y. Liu, Q. Liu and F. Qin (2014), A statistical method based on remote sensing for the estimation of air temperature in China. Int. J. Climatol, doi: 10.1002/joc.4113.

Dash, P., F. M. Goettsche, F. S. Olesen, H Fischer (2002), Land surface temperature and emissivity estimation from passive sensor data: Theory and practice-current trends, International Journal of Remote Sensing, 23(13), 2563-2594.Benali A, A. C. Carvalho, J. P. Nunes, N. Carvalhais, A. Santos (2012), Estimating air surface temperature in Portugal using MODIS LST data, Remote Sens. Environ. 124: 108–121, doi: 10.1016/j.rse.2012.04.024.

Duda, D. P. and P. Minnis (2000), A Study of Skin Temperature/Cloud Shadowing Relationships at the ARM SGP Site, Tenth ARM Science Team Meeting Proceedings, San Antonio, Texas, March 13-17, 2000.

Edwards J. M., J. R. McGregor, M. R. Bush, F. J. Bornemann (2011), Assessment of numerical weather forecasts against observations from Cardington: seasonal diurnal cycles of screen-level and surface temperatures and surface fluxes. Q. J. R. Meteorol. Soc. 137: 656–672. DOI:10.1002/qj.742.

Farr, T. G., and M. Kobrick (2000), Shuttle radar topography mission produces a wealth of data, Eos Trans. AGU, 81(48), 583–585, doi:10.1029/EO081i048p00583.

Freitas, S. C., I. F. Trigo, J. M. Bioucas-Dias, and F. M. Göttsche (2010), Quantifying the uncertainty of land surface temperature retrievals from SEVIRI/Meteosat, IEEE Trans. Geosci. Remote Sens., 48(1), 523–534.

Freitas, S. C., I. F. Trigo, J. Macedo, C. Barroso, R. Silva, and R. Perdigão (2013), Land surface temperature from multiple geostationary satellites, Int. J. Remote Sens., 34(9–10), 3051–3068, doi:10.1080/01431161.2012.716925.

Gallo, K., R. Hale, D. Tarpley, and Y. Yu (2011), Evaluation of the relationship between air and land surface temperature under clear- and cloudy-sky conditions. J. Appl. Meteorol. Climatol. 50, 767–775.

Good, E. (2015), Daily minimum and maximum surface air temperatures from geostationary satellite data, JGR-Atmospheres, doi:10.1002/2014JD022438.

Good, E (2016), 'An in situ-based analysis of the relationship between skin and screen-level temperatures', submitted to JGR-Atmospheres

Hachem S., S. C.R. Duguay , and M. Allard (2012), Comparison of MODIS-derived land surface temperatures with ground surface and air temperature measurements in continuous permafrost terrain, The Cryosphere, v6, 51–69, doi:10.5194/tc-6-51-2012.

Hall, D. K., V. V. Salomonson, and G. A. Riggs (2006), MODIS/Terra Snow Cover Daily L3 Global 0.05Deg CMG, Version 5. [MYD10.C1]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. http://dx.doi.org/10.5067/EI5HGLM2NNHN.

Haylock, M.R., N. Hofstra, A.M.G. Klein Tank, E.J. Klok, P.D. Jones and M. New (2008), A European daily high-resolution gridded dataset of surface temperature and precipitation. J. Geophys. Res (Atmospheres), 113, D20119, doi:10.1029/2008JD10201.

Janatian, N., Sadeghi, M., Sanaeinejad, S. H., Bakhshian, E., Farid, A., Hasheminia, S. M. and Ghazanfari, S. (2016), A statistical framework for estimating air temperature using MODIS land surface temperature data. Int. J. Climatol.. doi:10.1002/joc.4766

Jin, M., R. E. Dickinson and A. M. Vogelmann (1997), A Comparison of CCM2–BATS Skin Temperature and Surface-Air Temperature with Satellite and Surface Observations, Journal of Climate, Volume 10, pp 1505-1524.

Jin, M. and R. E Dickinson (2010), Land surface skin temperature climatology: benefitting from the strengths of satellite observations, Environ. Res. Lett., 5, 044004 doi:10.1088/1748-9326/5/4/044004.

Jin, M. S. and T. Mullens (2014), A Study of the Relations between Soil Moisture, Soil Temperatures and Surface Temperatures Using ARM Observations and Offline CLM4 Simulations, Climate 2014, 2, 279-295; doi:10.3390/cli2040279.

Kilibarda, M., T. Hengl, G. B. M. Heuvelink, B. Gräler, E. Pebesma, M. Perčec Tadić, and B. Bajat (2014), Spatio-temporal interpolation of daily temperatures for global land areas at 1km resolution, J. Geophys. Res. Atmos., 119, 2294–2313, doi:10.1002/2013JD020803.

Kilpatrick et al. (2015) A decade of sea surface temperature from MODIS, Remote Sensing of Environment, doi:10.1016/j.rse.2015.04.023.

Li, Z.-L., B.-H. Tang, H. Wu, H. Ren, G. Yan, Z. Wan, I.F. Trigo and J.A. Sobrino (2013), Satellite-derived land surface temperature: Current status and perspectives, Remote Sens. Environ., 131, 14–37, doi: 10.1016/j.rse.2012.12.008.

Lindgren, F., C. Morice, J. Kennedy (2015), 'Data calibration models and uncertainty specifications', report for the EUSTACE Horizon2020 project.

Mather, J. H. and J. W. Voyles (2013), The ARM Climate Research Facility: a review of structure and capabilities, Bulletin of the American Meteorological Society, 94(3), 377-392, doi:10.1175/BAMS-D-11-00218.1.

Menne, M.J., I. Durre, R.S. Vose, B.E. Gleason, and T.G. Houston (2012), An overview of the Global Historical Climatology Network Daily Database, Journal of Atmospheric and Oceanic Technology, DOI:http://dx.doi.org/10.1175/JTECH-D-11-00103.1Merchant, C.J. Ghent, D.J., Kennedy, J.J., Good, E. and Hoyer, J. (2015), Common approach to providing uncertainty estimates across all surfaces, EUSTACE Deliverable 1.2

Merchant, C.J.,Woolway, R.I., Ghent, D.J., and Hoyer, J. (2017), Uncertainties added to satellite datasets in CEMS, EUSTACE Deliverable 1.3

Mildrexler, D. J., M. Zhao, and S. W. Running (2011), A global comparison between station air temperatures and MODIS land surface temperatures reveals the cooling role of forests, J. Geophys. Res., 116, G03025, doi:10.1029/2010JG001486.

Oyler, J. W., A. Ballantyne, K. Jencso, M. Sweet and S. W. Running (2015), Creating a topoclimatic daily air temperature dataset for the conterminous United States using homogenized station data and remotely sensed land skin temperature. Int. J. Climatol, 35, 2258–2279. doi:10.1002/joc.4127.

Parmentier, B., B. J., A. M. McGill, Wilson, J. Regetz, W. Jetz,, R. Guralnick, M.-N. Tuanmu, and M. Schildhauer (2015), Using multi-timescale methods and satellite-

derived land surface temperature for the interpolation of daily maximum air temperature in Oregon, Int. J. Climatol., 35, 3862–3878, doi:10.1002/joc.4251.

Poulter, B., N. MacBean, A. J. Hartley, I. Khlystova, O. Arino, R. A. Betts, S. Bontemps, M. Boettcher, C. Brockmann, P. Defourny, S. Hagemann, M. Herold, G. Kirches, C. Lamarche, D. Lederer, C. Ottlé, M. Peters, and P. Peylin (2015), Plant functional type classification for Earth System Models: results from the European Space Agency's Land Cover Climate Change Initiative. Geosci. Model Dev. Discuss. 8(1), 429–462 (doi:10.5194/gmdd-8-429-2015)

Rodriguez, E., C. S. Morris, J. E. Belz, E. C. Chapin, J. M. Martin, W. Daffer, S. Hensley (2005), An assessment of the SRTM topographic products, Technical Report JPL D-31639, Jet Propulsion Laboratory.

Stokes, G. M., and S. E. Schwartz (1994), The Atmospheric Radiation Measurement (ARM) Program: Programmatic Background and Design of the Cloud and Radiation Test Bed. Bull. Amer. Meteor. Soc., 75, 1201–1221. doi: http://dx.doi.org/10.1175/1520-0477(1994)075<1201:TARMPP>2.0.CO;2.

Sohrabinia, M. , P. Zawar-Reza and W. Rack (2014), Spatio-temporal analysis of the relationship between LST from MODIS and air temperature in New Zealand, Theor. Appl. Climatol., DOI 10.1007/s00704-014-1106-2.

Trigo, I. F., I. T. Monteiro, F. Olesen, and E. Kabsch (2008), An assessment of remotely sensed land surface temperature, J. Geophys. Res., 113, D17108, doi:10.1029/2008JD010035.

Urban, M., J. E., C. Hüttich, C. Schmullius, and M. Herold (2013), Comparison of Satellite-Derived Land Surface Temperature and Air Temperature from Meteorological Stations on the Pan-Arctic Scale, Remote Sens. 2013, 5, 2348-2367; doi:10.3390/rs5052348.

Vancutsem, C., P. Ceccato, T. Dinku, S. J. Connor (2010), Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa, Remote Sens Environ, 114(2), 449–465, doi:10.1016/j.rse.2009.10.002, http://www.sciencedirect.com/science/article/pii/S0034425709003113.

Wan, Z. (2013), Collection 6 MODIS Land Surface Temperature Products Users' Guide, http://www.icess.ucsb.edu/modis/LstUsrGuide/MODIS_LST_products_Users_guide_Collection-6.pdf.

Wan, Z. (2014), New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product, Remote Sens Environ, 140, 36-45.

3.4 Relationships over ice

3.4.1 Summary of ice relationships

The Polar Regions have a sparse in situ observational network, due to the extreme conditions and the low population. The potential for using satellite observations of ice surface temperature (IST) to estimate the near surface air temperatures (IAT) is therefore large in these regions. IST and IAT are often well coupled, and the two temperatures can be close (within 1°C) during certain conditions, e.g. when clouds are present or during strong winds. However, IST and IAT frequently differ by several degrees or more, with the largest differences occurring during night and under cloud-free skies.

In this study, we estimate daily IAT on the basis of Earth Observation satellite retrievals of skin temperatures, using a linear regression model. In the optimal regression model, IAT is predicted from daily satellite IST combined with a seasonal variation. The new surface air temperature data record has been validated against independent, in situ measured near surface air temperatures. Moreover, we provide a direct comparison of the air temperature data record against the existing near surface air temperature reanalysis product, ERA-Interim.

3.4.2 Data

3.4.2.1 In situ data

In situ observations have been collected from various sources and campaigns covering ice and snow surfaces to augment an existing DMI database. The DMI database includes quality controlled and formatted observations from 2,910 stations/buoys/ships with temperature observations covering ice and snow surfaces (Figure 3.4.1). Here, we use a subset of temperature observations from the DMI database covering the period 2000-2009. Table 3.4.1 shows a summary of the 329 sites/buoys/ships used here. For the Northern

Hemisphere land ice we use data from PROMICE[1], ARM[2] data from Alaska, and GC-Net[3] data. The northern sea ice is covered mainly by data obtained from the ECMWF operational data stream, providing approx. 200 unique data series from drifting buoys used in this study. A few data series included here provide both surface and air temperatures across the Northern Hemisphere sea ice. These include the Sea Ice Temperature and Radiation Data set provided by DMI, the SHEBA[4] data and MET Norway IST temperature and radiative flux measurements[5]. Fewer in situ observations are available for the Southern Hemisphere sea ice and a total of 13 drifting buoys are used from the ECMWF operational data stream. The Southern Hemisphere land ice is covered by NCEI[6], AMRC[7] and one data set from Australian Antarctic Data Centre (AADC) automatic weather stations (AWS).
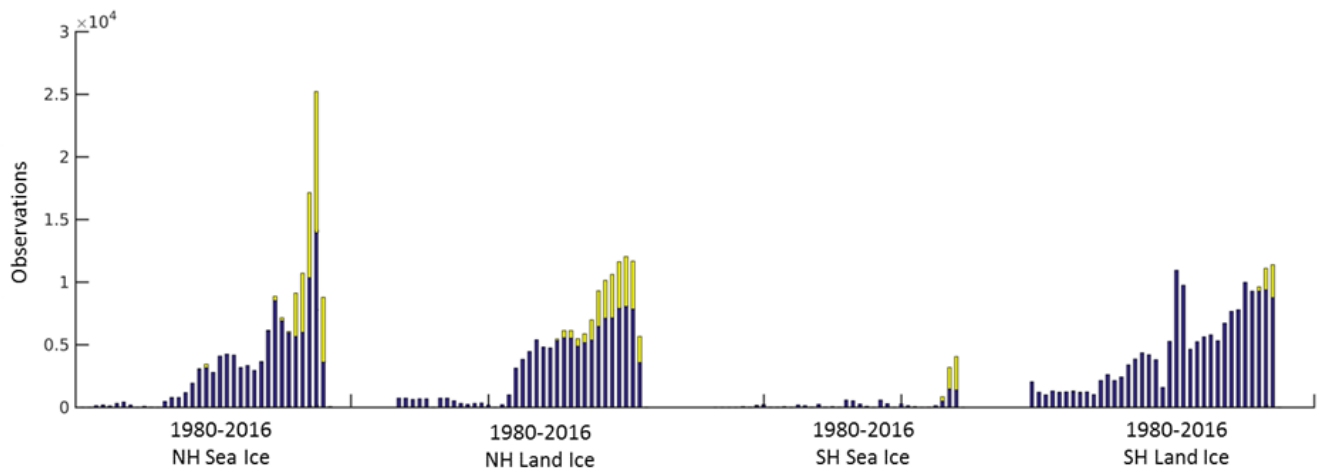

Figure 3.4.1: *Total number of observations in the period 1980-2016. The portion of observations which are of near surface air temperature is shown in blue and the portion of observations which are of ice surface temperature is shown in yellow.*

---

[1] Programme for Monitoring of the Greenland Ice Sheet by the Geological Survey of Denmark and Greenland (GEUS), http://www.promice.dk/home.html

[2] Atmospheric Radiation Measurement Program, http://www.arm.gov/sites/nsa

[3] Greenland Climate Network by Cooperative Institute for Research in Environmental Sciences (CIRES), http://cires1.colorado.edu/steffen/gcnet/

[4] Surface Heat Budget of the Arctic from the Canadian DesGroseilliers icebreaker http://www.eol.ucar.edu/projects/sheba

[5] https://sabvabaa.nersc.no/node/395

[6] National Centers for Environmental Information, formerly the National Climatic Data Center (NCDC), ftp://sidads.colorado.edu/pub/DATASETS/surface_obs/

[7] Antarctic Meteorological Research Center, https://amrc.ssec.wisc.edu/aws/

Uncertainties of the temperatures measured *in situ* have been estimated based on information on the observational platform and by comparison with the satellite data. A fixed uncertainty estimate has been assigned to data from each data provider and surface type, ranging from 0.5°C to 2.2°C (Table 3.4.1).

Table 3.4.1: *Overview of in situ data for 2000-2009 in the DMI database used and their assigned uncertainties.*

	No. of sites, buoys or cruises	No. of daily observations	Spatial Coverage	Surface Type	Observation Type	Parameters	Assigned instrument uncertainty (°C)
AADC	1	818	SH	Land ice	AWS	T2m	0.5
ACSYS	7	280	NH	Sea ice	Buoy	T2m	1.0
AMRC	25	31,912	SH	Land ice	AWS	T2m, Tskin	0.5
ARM	2	2,846	NH	Land snow	AWS	T2m, Tskin	1.0
CRREL	10	1,031	NH	Sea ice	Buoy	T2m	0.5
DAMOCLES	25	2,160	NH	Sea ice	Buoy	T2m	1.0
ECMWF	209	27,235 NH, 840 SH	196 NH, 13 SH	Sea ice	Buoy	T2m, Tskin	1.5
FRAMZY	11	251	NH	Sea ice	Buoy	T2m	1.0
GC-NET	15	29,133	NH	Land ice	AWS	T2m, Tskin	2.2
NCDC	14	3,732	SH	Land ice	AWS	T2m, Tskin	1.5
Polarstern	2	189 NH, 210 SH	1 NH, 1 SH	Sea ice	Ship	T2m	0.5
PROMICE	8	2,685	NH	Land ice	AWS	T2m, Tskin	0.5

PROMICE has proven very useful in this analysis and was also used earlier within EUSTACE as a reference data set for validation of satellite IST, fitting different regression models, and validation of the final IAT data product. PROMICE delivers in situ observations from a number of AWSs along the margin of the Greenland Ice Sheet. Only high-altitude stations have been used in this study in order to ensure a permanent snow/ice covered surface. PROMICE IST has been calculated from the up-welling longwave radiation assuming blackbody radiation properties (Ahlstrøm et al. 2008).

3.4.2.2 Satellite data

The satellite data used in this study are from the Arctic and Antarctic Ice Surface Temperatures from thermal Infrared satellite sensors (AASTI; Dybkjær et al., 2014) data set,

covering high latitude seas, sea ice, and ice cap surface temperatures based on satellite infrared measurements (Karlsson et al., 2013). The data set is based on one of the longest existing satellite records from the Advanced Very High Resolution Radiometer (AVHRR) instruments aboard a long series of NOAA satellites. The first version of the AASTI product, used here, is available from 2000 to 2009 in approximately 5 km spatial resolution. The different AVHRR satellites have been orbiting 14 times per day, providing approximately bi-hourly coverage of the Polar Regions. An increasing number of satellites has been in orbit from 2000 to 2009 (Figure 3.4.2).

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Start date	End date
NOAA14											1995/01/20	2001/12/31
NOAA15											2000/01/01	
NOAA16											2001/01/01	
NOAA17											2002/06/25	
NOAA18											2005/05/20	
METOP2											2007/06/28	
NOAA19											2009/02/06	

Figure 3.4.2: *NOAA and Metop satellites carrying the AVHRR sensor, used for AASTI version 1.*

The level 2 satellite observations were aggregated into daily level 3 averages of ice surface temperatures on a fixed 0.25 by 0.25 degrees grid, which is common for all surfaces. EUSTACE is developing products for days defined by local solar time. The data in the aggregated files thus contains observations from 00 to 24 local solar time. The main aggregated surface variables are the daily average, minimum and maximum surface temperatures, but 3-hourly averages of surface temperatures and the number of observations were also calculated for the eight intervals during each day. These 3-hourly averages were used for estimating the satellite sampling throughout the day and to gain confidence in the daily cycle estimates. In the aggregation, all satellite observations with a quality flag of 4 or 5 were used. The aggregated surface skin data set is available to users upon request. Figures 3.4.3 and 3.4.4 show the mean number of observations per day in each of the eight time intervals given in local time for the Northern and Southern Hemisphere, respectively. The variation in the coverage throughout the day is a combined effect of the satellite overpasses and the availability of cloud-free conditions during the day.

The cloud mask in the PROMICE data set displays a minimum in the cloud coverage around noon, local time (not shown), which is consistent with Figure 3.4.3.
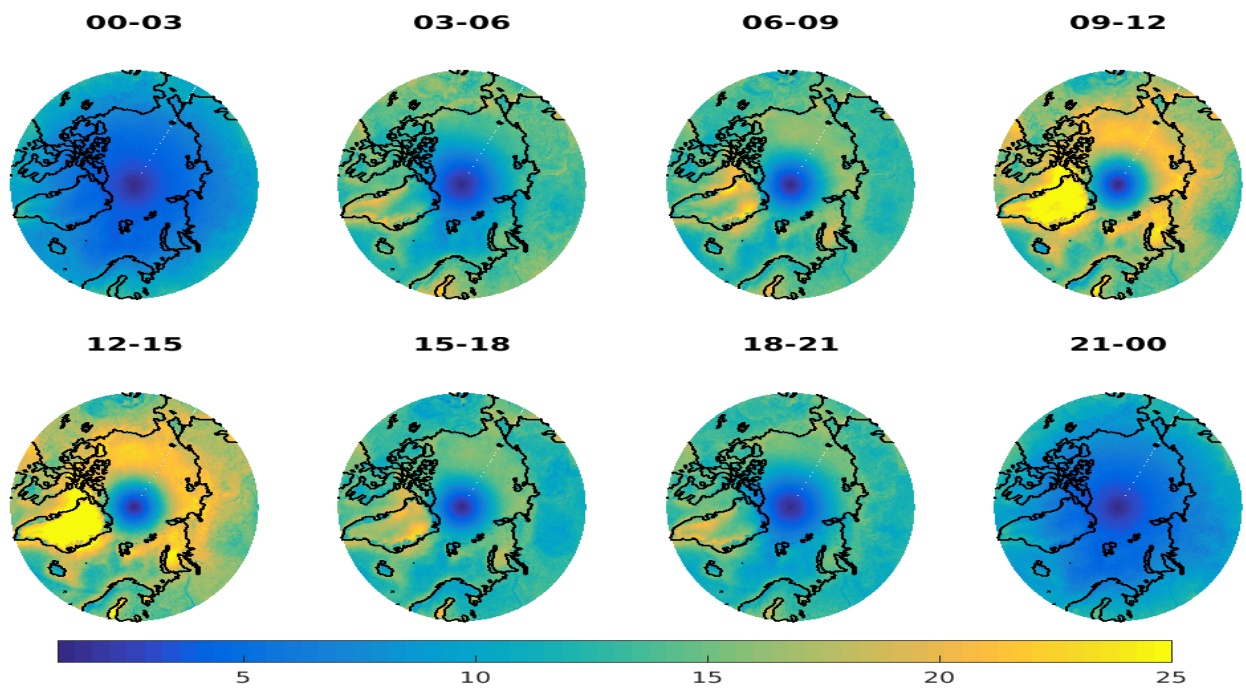

Figure 3.4.3: *Mean number of observations per day in each of the eight local solar time intervals. Northern Hemisphere averages 2000-2009.*
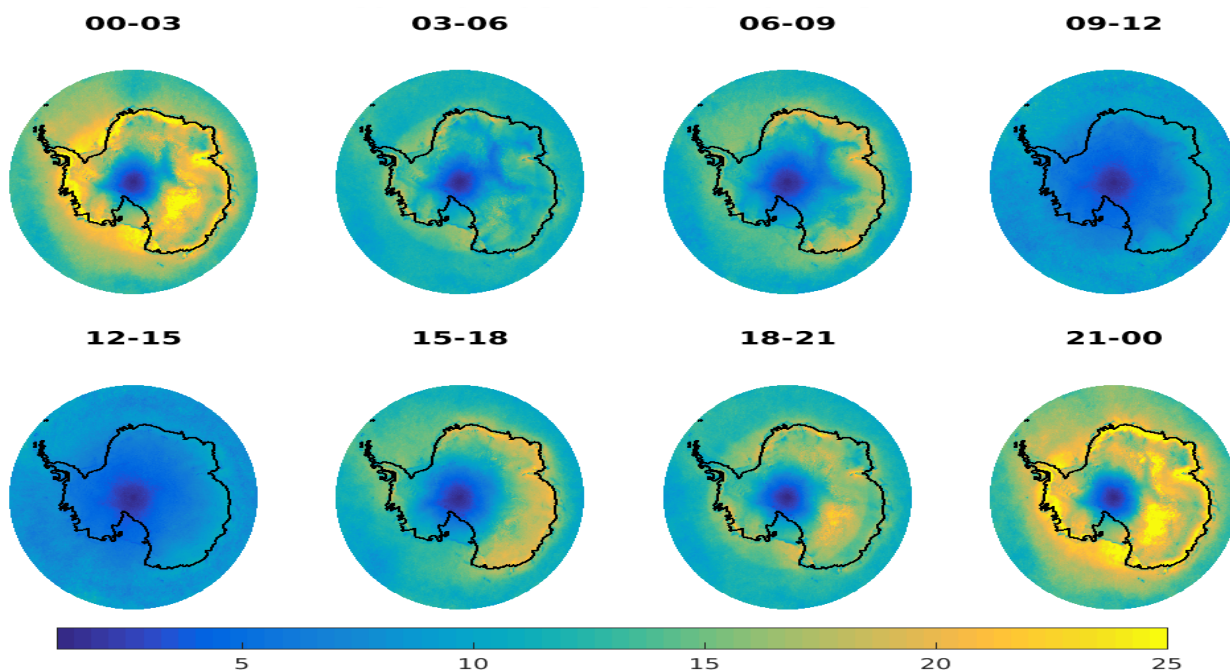
Figure 3.4.4: *Mean number of observations per day in each of the eight local solar time intervals. Southern Hemisphere averages 2000-2009.*

In order to best resolve the diurnal cycle with satellite information without losing too many observations we require data during both night (6pm-6am) and day (6am-6pm) in order to calculate the daily mean temperature (Tmean). Here, Tmean is calculated by averaging all available observations during the day. To calculate the daily minimum (Tmin) we require observations in at least one of the night bins and for daily maximum (Tmax) we require observations in at least one of the day bins. A few more criteria have been set up in order to limit temporal sampling errors, effects of undetected clouds and outliers:

- Tmean, Tmin and Tmax are not allowed to exceed +5°C, to avoid clear melting conditions and obviously wrong measurements.
- If the standard deviation of satellite IST during one day exceeds 7.07°C, corresponding to a sinusoidal daily cycle with a difference between day and night of 20°C, no Tmean, Tmin or Tmax are calculated.
- If the difference between Tmean and the average of all available averaged bin temperatures exceeds 10°C, Tmean is discarded.

- If Tmin is more than 20°C colder than the night bins average, Tmin is left out, because this indicates cloud contamination of the Tmin value.
- If Tmax is more than 20°C warmer than the day bins average, Tmax is left out, because this indicates an outlier, possibly from an open water patch in sea ice.
- If Tmean/Tmax/Tmin is more than 10°C colder than the corresponding average of up to 24 neighbouring cells with same surface type, Tmean/Tmin/Tmax is left out.

3.4.2.3 Validation of satellite data

AASTI v.1 Level 3 data has been validated against in situ observations from PROMICE and ARM stations covering land ice in the Northern Hemisphere. Table 3.4.2 shows the results of AASTI v.1 Level 3 data (IST_{SAT}) validated against in situ ice surface temperature measurements (IST_{InSitu}) and near surface air temperature measurements (IAT_{InSitu}), respectively. In general, IST_{SAT} correlates better with the in situ measured air temperature than with the in situ measured ice surface temperature. Moreover, the IST_{SAT}-IAT_{InSitu} difference shows a smaller standard deviation than IST_{SAT}-IST_{InSitu}. However, as expected the bias and RMS are larger for IST_{SAT}-IAT_{InSitu} difference for the ten weather stations included in this analysis.

Table 3.4.2: *Validation of AASTI v.1 Level 3 IST against in situ IST and IAT observations.*

Station	N	ISTsat – ISTinsitu				ISTsat - IATinsitu			
		corr	Bias (°C)	std (°C)	RMS (°C)	corr	Bias (°C)	std (°C)	RMS (°C)
ARM_Atq	1235	93.8	-2.47	3.69	4.44	93.7	-3.17	3.69	4.87
ARM_Bar	1594	94.1	-0.73	4.30	4.36	94.6	-1.14	4.02	3.86
PROMICE KAN-M	422	93.9	-3.65	3.37	4.96	94.6	-4.56	3.14	5.53
PROMICE KAN-U	239	93.9	-1.75	3.32	3.75	94.4	-3.39	3.17	4.64
PROMICE KPC-U	488	97.6	-1.31	2.62	2.92	98.2	-3.20	2.27	3.92
PROMICE NUK-U	296	77.7	-4.09	5.00	6.45	84.7	-7.19	4.01	8.23
PROMICE QAS-U	407	83.9	-1.65	4.20	4.51	86.3	-3.70	3.75	5.27
PROMICE SCO-U	403	91.5	-4.60	4.25	6.26	93.7	-7.55	3.75	8.43
PROMICE TAS-U	386	67.5	-1.03	5.43	5.52	79.5	-3.61	4.39	5.68
PROMICE UPE-U	125	88.2	-3.13	3.88	4.97	90.0	-5.49	3.50	6.50
All data	**5595**	**92.9**	**-2.03**	**4.24**	**4.70**	**93.2**	**-3.36**	**4.12**	**5.32**

3.4.3 Variability of the IST vs air temperature relationships

3.4.3.1 Variability from in situ

Here we perform an analysis of the IST and IAT variability and the factors controlling the surface and air temperature variations in ice and snow covered regions. The IST-IAT

relationship has been studied using *in situ* measurements from ten PROMICE AWSs covering the margin of the Greenland Ice Sheet. One example, for the north-western station in Upernavik, is shown in Figure 3.4.5. Here, three different sky conditions have been compared in terms of IAT-IST differences as a function of local time and season. In the case of all-sky conditions the winter diurnal cycles display very little variability, as expected from the lack of insolation. The vertical stratification is weakest in spring and fall around noon and early afternoon, when the sun is out and can warm the surface. At these times, IST may even exceed IAT slightly. During night, the surface is often colder than the atmosphere resulting in a strong vertical stratification.

The difference in downwelling longwave radiation between clear-sky and overcast conditions can result in changes in the differences between IAT and IST of up to around 5°C (Figure 3.4.5).
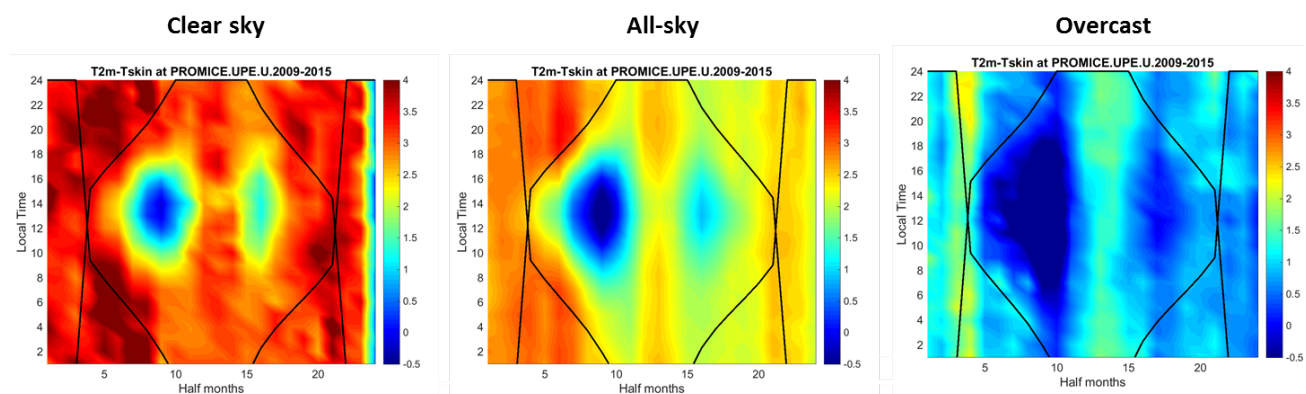

Figure 3.4.5: *Observed IAT (T2m) minus IST (Tskin) differences (°C) at the upper station in Upernavik in cases of clear-sky, all-sky, and overcast conditions. The black lines indicate times of sun rise and sun set.*

During clear skies, large differences (>3°C) between IAT and IST are prevalent most of the time. Overcast conditions lead to a weaker stratification and, in some cases, a reversal with very limited diurnal variability. All PROMICE stations included in this study show that the strongest stratification occurs during clear-sky conditions. This must be taken into account when deriving IAT from IST, as infrared satellite observations are only available during clear skies.

Figure 3.4.6 shows the impact of wind on the IAT-IST differences for the upper PROMICE station in Upernavik. For all stations, the strongest vertical stratification occurs at wind speeds of the order of 5 m/s. Higher wind speeds force turbulent mixing and a weak vertical temperature gradient between IST and IAT. At wind speeds weaker than 5 m/s the vertical stratification is vanishing and may be reversed. In general, though, the variance in IAT-IST difference decreases with increasing wind speed.
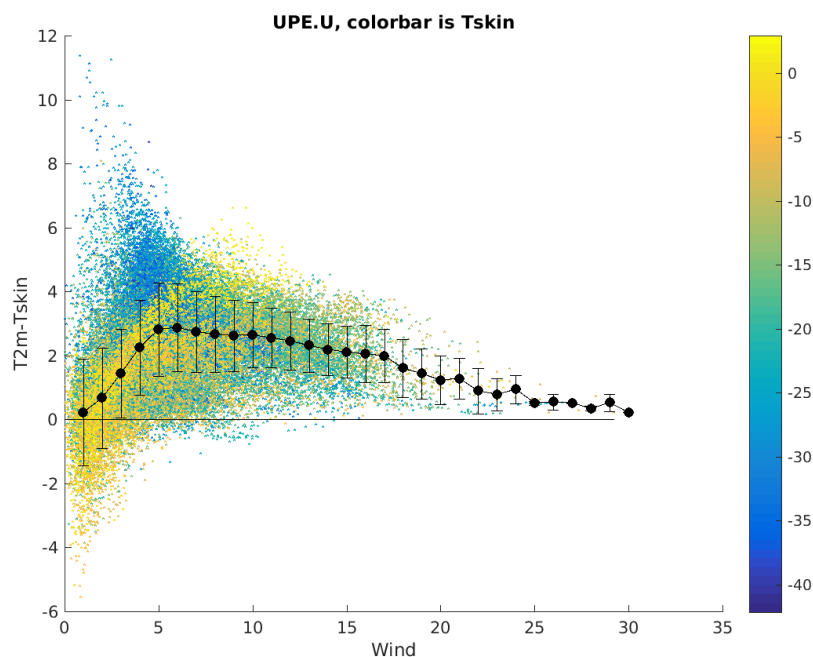

Figure 3.4.6: *Scatterplot showing the observed IAT (T2m) minus IST (Tskin) differences (°C) at the upper station in Upernavik as a function of wind and IST. The black lines display the mean temperature difference and the corresponding error bars. The colours represent the IST.*

3.4.3.2 Variability from Satellite

The satellite-derived IST shows a significant seasonal difference in daily variability, with largest standard deviations during summer in East Antarctica and during spring for Southern Hemisphere sea ice (Figure 3.4.7). The freeze-up of sea ice also causes higher variability along the sea ice margin.
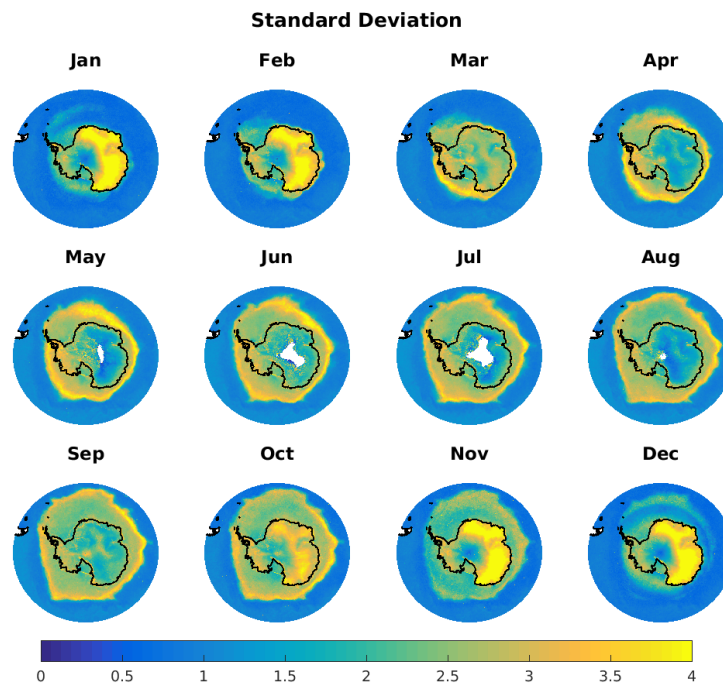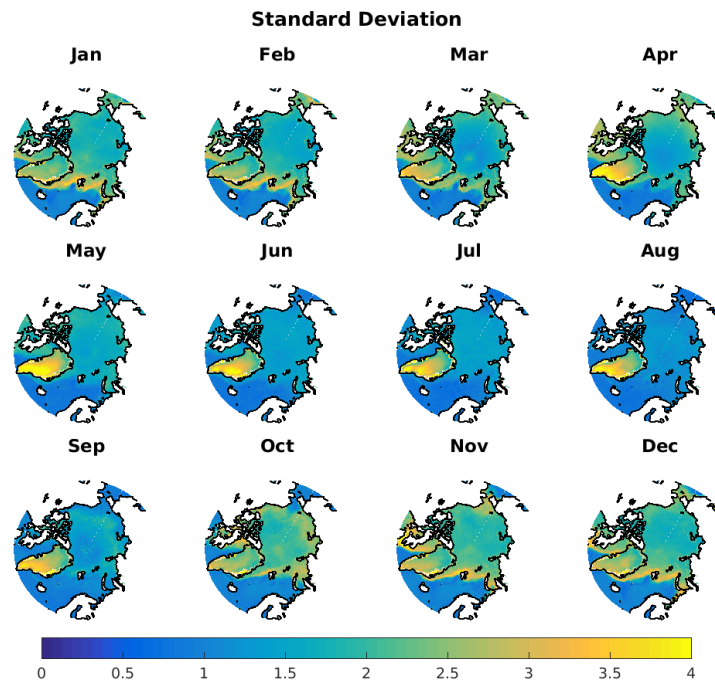
Standard Deviation

Standard Deviation

Figure 3.4.7: *Standard deviations (°C) of daily satellite surface temperature observations by month for the Northern (top) and Southern (bottom) Hemisphere. Averages over years 2000-2009.*

3.4.4 Relationship model and application to satellite data

A regression-based approach has been used to estimate daily Tmean, Tmin and Tmax at the grid scale using AASTI satellite data, providing spatially-detailed observations for areas unobserved by in situ stations.

3.4.4.1 Matchup

The AASTI data have been matched with in situ data for each day by requiring a distance to nearest in situ station of less than 0.5 degree (approx. 55km in latitude). Separate regression models are produced for land and sea ice in both hemispheres. The match-up stations have been divided into two sub-sets for each domain; one set used to generate regression coefficients and another set used to validate the regression model (Figure 3.4.8).
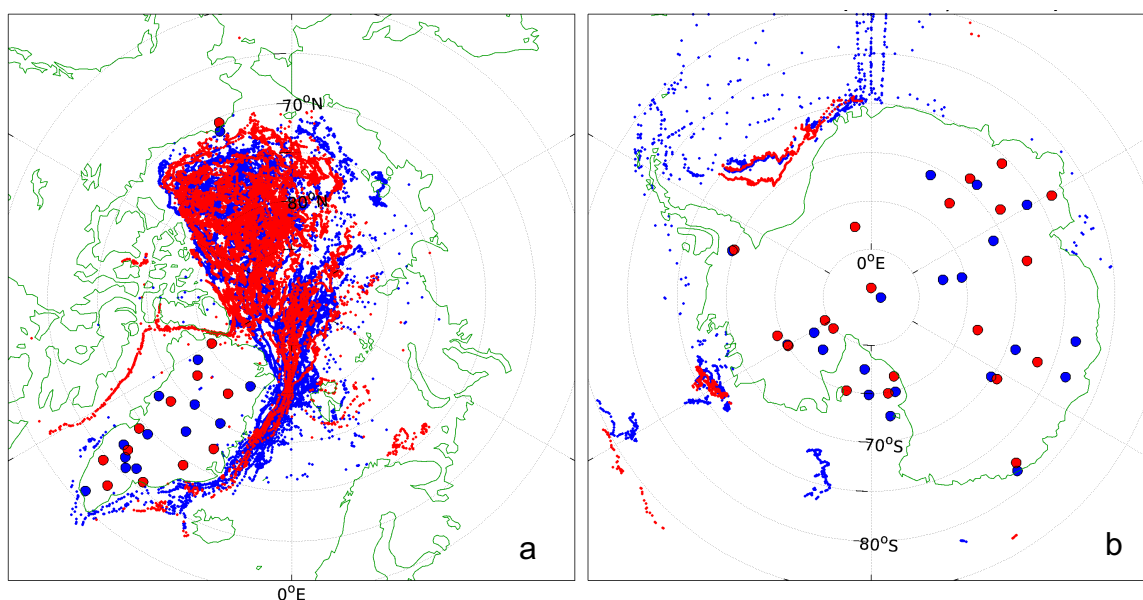

Figure 3.4.8: *Positions of match-ups on sea ice and land ice in the Northern Hemisphere (a) and Southern Hemisphere (b) (red: analysis/training, blue: validation).*

3.4.4.2 Regression model

The regression model is based on multiple linear regression analysis using least squares (Menke 1989). As explaining factors, latitude, downward shortwave radiation not considering clouds (theoretical), seasonal cycle, and wind (ERA-Interim reanalysis) were tested. These factors were selected for testing based on current knowledge from the literature and an analysis of parameters that influence the relationship between IST and IAT at PROMICE stations and field experiments in Qaanaaq, limited by the available data.

The best correlation against training data was found using a model where IAT is predicted from daily satellite IST combined with a seasonal variation assumed to have the shape of a cosine function $A \cdot cos\left(\frac{time \cdot 2\pi}{1\ year} - \varphi\right)$ where A is the amplitude and φ is the phase. This can be rewritten in the form:

$$T_{mean} = \alpha_0 + \alpha_1 \cdot IST_{avg} + \alpha_2 \cdot cos\left(\frac{time \cdot 2\pi}{1\ year}\right) + \alpha_3 \cdot sin\left(\frac{time \cdot 2\pi}{1\ year}\right) + \varepsilon \qquad (Eq.\ 3.4.1)$$

suitable for linear regression. The training data have been used to calculate the regression coefficients for each regression model pertaining to land and sea ice in both hemispheres (Table 3.4.3), giving a model fit with correlation of 94-96% and standard deviations of 3.2-3.3°C against training data for all four ice surfaces (Table 3.4.4). Similar regression coefficients have been derived for Tmin ($\beta_0\ to\ \beta_3$) and Tmax ($\gamma_0\ to\ \gamma_3$).

Table 3.4.3: *Model regression coefficients*

		Offset (°C) α_0	IST factor α_1	Cos amplitude (°C) α_2	Sin amplitude (°C) α_3
Land Ice	Northern Hemisphere	4.20	1.06	2.14	-0.74
	Southern Hemisphere	5.70	1.04	-0.42	-0.22
Sea Ice	Northern Hemisphere	1.46	0.89	-1.34	-1.24
	Southern Hemisphere	1.41	0.87	0.96	0.76

Table 3.4.4: *Statistics on the relation between observed and modelled IAT for the training data*

	Number of observations	Correlation (%)	Standard deviation (°C)	Minimum difference (°C)	Maximum difference (°C)
Land ice N Hemisphere	13792	96.3	3.28	-12.9	12.2
Land ice S Hemisphere	15122	96.5	3.26	-10.9	9.9
Sea ice N Hemisphere	15010	96.2	3.25	-11.4	11.9
Sea ice S Hemisphere	430	94.4	3.16	-11.5	8.6

3.4.4.3 Estimating air temperature

The derived coefficients have been used to estimate IAT over land and sea ice from satellite IST during the period 2000-2009. Four datasets are developed covering the Northern Hemisphere sea ice, Northern Hemisphere land ice, Southern Hemisphere sea ice, and Southern Hemisphere land ice. Each dataset provides daily estimates of minimum air temperature (Tmin), maximum air temperature (Tmax), and mean air temperature (Tmean). The data sets are stored in compressed NetCDF-4 format and contain global IAT estimates on a 0.25 degree regular latitude-longitude grid. Each temperature estimate is associated with three categories of uncertainty: random uncertainties on the 0.25 degree daily scale; synoptic scale correlated uncertainty; and globally correlated uncertainty, including and excluding uncertainties related to the masking of clouds. The three types of uncertainties are also gathered in two total uncertainty estimates: one which includes the cloud mask uncertainty, and one which does not.

Figure 3.4.9 shows the monthly mean near surface air temperature estimates averaged over the Greenland Ice Sheet (see also Figure 3.4.10 for an example map). The Greenland Ice Sheet exhibits a distinct annual cycle in near surface air temperature. The monthly mean air temperatures typically reach a maximum of -4°C during July and a minimum of about -28°C during winter. As is common for Arctic climates, temporal variability is largest during winter due to a more vigorous atmospheric circulation.
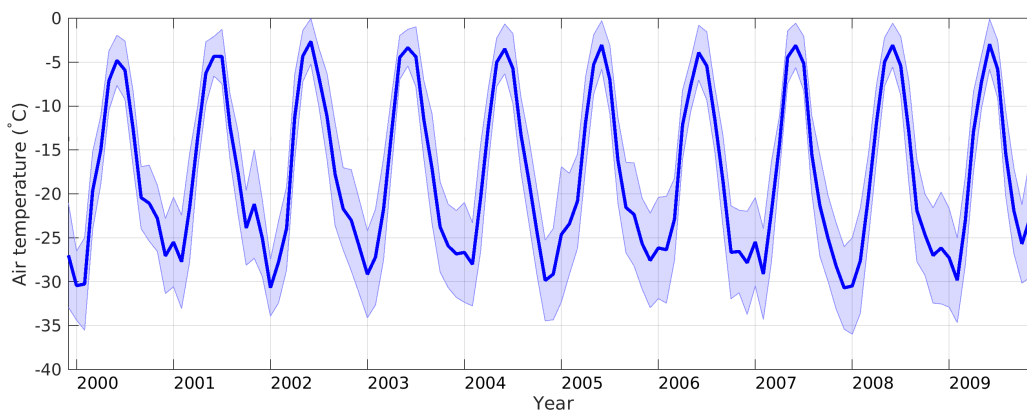

Figure 3.4.9: *Monthly mean satellite-derived near surface air temperature estimates (°C) for the Greenland Ice Sheet. The shading represents the variability over the month.*

3.4.5 Modelling uncertainties

Uncertainty estimates for the satellite-derived IAT data are split into random, locally correlated and systematic uncertainties. The implementation for a particular grid cell at a particular point in time is outlined below.

3.4.5.1 Satellite IST uncertainties

First, we propagate the different uncertainty components for the IST retrievals onto the 0.25° latitude by 0.25° longitude grid used here.

The random uncertainty component of the gridded satellite IST data is given by:

$$RU_{sat} = \sqrt{RU_{instrument}^2 + RU_{geolocation}^2} \qquad \text{(Eq. 3.4.2)}$$

where $RU_{instrument}$ and $RU_{geolocation}$ are the uncertainty components due to instrument noise and geolocation errors in the marginal ice zone, respectively, from the same grid cell at the same point in time.

The locally correlated uncertainties are given by:

$$SSU_{sat} = \sqrt{SSU_{emissivity}^2 + SSU_{atmosphere}^2} \qquad \text{(Eq. 3.4.3)}$$

where $SSU_{emissivity}$ and $SSU_{atmosphere}$ are the synoptic scale uncertainty components of the IST data due emissivity errors and atmospheric corrections.

The grid-cell systematic uncertainties (LSU_{sat}) are based on expert judgements and provided as a fixed value of 0.2°C.

The cloud mask uncertainty is calculated from the cloud mask quality level of the grid cell, using

$$CU_{sat,avg} = \alpha_1 \cdot (\delta_{avg} + 0.5 \cdot (5 - QL)) \qquad \text{(Eq. 3.4.4)}$$

where α_1 is the regression coefficient (see Table 3.4.2) and δ_{avg} is an overall uncertainty level for CU_{avg}, set to 0.8°C. The uncertainty components for Tmin and Tmax are found by replacing α_1 with β_1 and γ_1 and δ_{avg} with δ_{min}=2.8°C and δ_{max}=0.8°C. The higher δ-value for Tmin occurs because undetected clouds will appear as artificial cold temperatures,

especially affecting the Tmin field because they are harder to distinguish from the coldest surface temperatures.

3.4.5.2 IAT uncertainties

The random uncertainty component for Tmean is given by:

$$RU_{avg} = \sqrt{(\alpha_1 \cdot RU_{sat})^2 + RU_{sampling}^2} \qquad \text{(Eq. 3.4.5)}$$

where RU$_{sampling}$ is the uncertainty relating to sampling errors in space and time (see section 3.4.5.3). The random uncertainty components for Tmin and Tmax are found by replacing α_1 with β_1 and γ_1, respectively.

Similarly, the locally correlated (synoptic scale) uncertainty component is given by:

$$SSU_{avg} = \sqrt{(\alpha_1 \cdot SSU_{sat})^2 + SSU_{relation}^2} \qquad \text{(Eq. 3.4.6)}$$

The term SSU$_{relation}$ represents the standard deviation of the residuals calculated at in situ stations where both skin and air temperatures are available, i.e. the skin temperature-predicted Tavg minus the observed Tmean, calculated during the model training process of these in situ only data (Table 3.4.6). Again, the uncertainty components for Tmin and Tmax are found by replacing α_1 with β_1 and γ_1, respectively.

Table 3.4.6: *Uncertainty estimates for the relationship error*

	T$_{mean}$ (°C)	T$_{min}$ (°C)	T$_{max}$ (°C)
Land ice	1.5	1.8	2.0
Sea ice	1.7	1.8	2.1

Three forms of systematic uncertainties are provided: the covariance matrix of the model coefficients, an estimate of the systematic uncertainty in IST for each grid cell and the cloud mask uncertainty component. The grid-cell systematic uncertainties are $LSU_{avg} = \alpha_1 \cdot LSU_{sat}$ and cloud mask uncertainty $CU_{avg} = \alpha_1 \cdot CU_{sat}$, and as above, the uncertainty components for Tmin and Tmax are found by replacing α_1 with β_1 and γ_1, respectively.

3.4.5.3 Sampling uncertainty

The sampling uncertainty is determined indirectly by closing the average RMS error budget when validating against the independent in situ air temperature measurements, taking the in situ air temperature uncertainty into account. This approach was used due to limited time and limited validation data, but has the disadvantage that other unaccounted uncertainty sources are artificially grouped with the sampling uncertainty. The sampling uncertainty is determined individually for the average, minimum and maximum temperatures, and for each surface type, see Table 3.4.6.

Table 3.4.6: *Sampling uncertainty (°C) for Tmean, Tmin and Tmax.*

	T_{mean} (°C)	T_{min} (°C)	T_{max} (°C)
Land ice Northern Hemisphere	1.6	2.2	2.1
Sea ice Northern Hemisphere	0.0[8]	2.9	2.0
Land ice Southern Hemisphere	1.6	1.5	3.8
Sea ice Southern Hemisphere	1.7	5.7	2.9

3.4.6 Validation

3.4.6.1 Independent station evaluation

An evaluation of the IAT regression model performance has been carried out by comparison to the independent *in situ* data (see Figure 3.4.8). Figure 3.4.10 shows an example of daily near surface air temperatures from 01-01-2008 in both hemispheres. Circles are *in situ* near surface air temperatures from coincidence independent stations and buoys. The overall model performance when compared to all independent stations/buoys/ships is summarized in Table 3.4.7. Note that the uncertainty of the *in situ* data used for comparison is also accounted for (Section 3.4.2.1) and reflected in the results below.

---

[8] This zero-value is likely caused by a slight overestimation of some of the other uncertainty quantities.
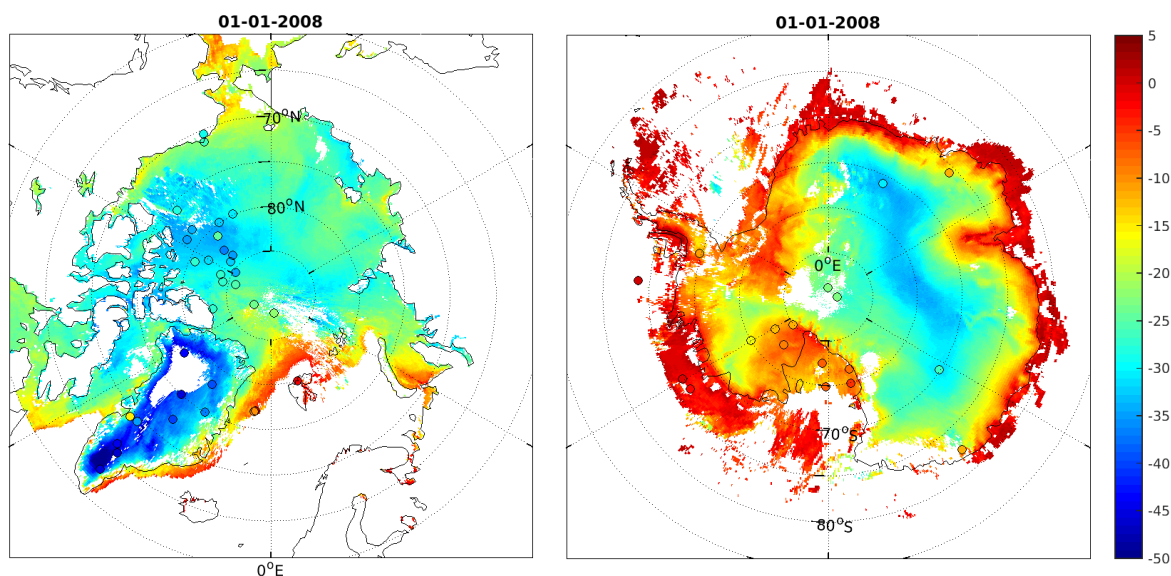
Figure 3.4.10: *Daily mean air surface temperature (°C) over land ice and sea ice from January 1, 2008. Circles show in situ measurements.*

Table 3.4.7: *Model performance (satellite-derived minus in situ measured air temperatures) when compared to all independent stations/buoys/ships.*

	Number of observations	Correlation (%)	Bias (satellite derived – in situ, °C)	Standard deviation (°C)	Root mean square difference (°C)
Land ice Northern Hemisphere	20872	95.5	0.30	3.45	3.47
Sea ice Northern Hemisphere	16092	96.5	0.35	3.18	3.19
Land ice Southern Hemisphere	21327	97.3	0.12	3.11	3.11
Sea ice Southern Hemisphere	620	91.9	0.20	3.55	3.55

3.4.6.2 Comparison with ERA-Interim

ERA-Interim (Dee et al., 2011) is a global atmospheric dynamical reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) covering the period from 1989 to present. This analysis compares the ability of ERA-Interim and the proposed regression model to reproduce near surface air temperatures at all independent stations (See Table 3.4.8).

Table 3.4.8: *Statistics on the relation between ERA-Interim and in situ measured temperatures for independent test data.*

	Number of observations	Correlation (%)	Bias (ERA Interim – in situ, °C)	Standard deviation (°C)	Root mean square difference (°C)
Land ice Northern Hemisphere	20872	96.4	3.41	3.18	4.66
Sea ice Northern Hemisphere	16092	96.9	1.14	3.02	3.22
Land ice Southern Hemisphere	21327	97.5	0.96	3.00	3.15
Sea ice Southern Hemisphere	620	95,8	0.38	2.50	2.52

The comparison may not be truly independent here as a number of stations/buoys have been assimilated into the ERA-Interim data product. The figures below give an indication of the performance of ERA-Interim and the satellite-derived IAT at individual stations/buoys/ships for each surface type and hemisphere.
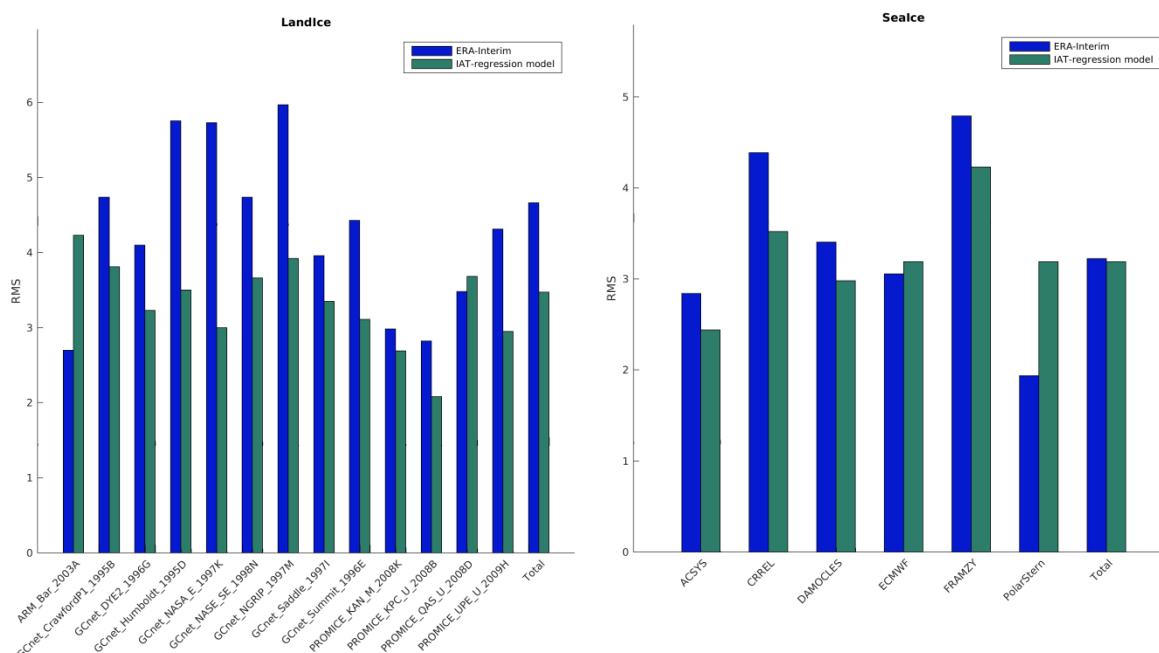

Figure 3.4.11: *Root mean square differences (°C) calculated for the Northern Hemisphere land ice (left) and sea ice (right). Blue columns are RMS obtained by comparing in situ with ERA-Interim while green columns are in situ compared with the satellite-derived IAT data*

set. Only buoys with more than 200 observations are included in the figure. The last columns listed as "total" are the RMS obtained by using all test data.

Figure 3.4.11 shows the RMS difference calculated between the independent test stations and ERA-Interim and between independent test stations and the satellite-derived air temperatures for both surface types in the Northern Hemisphere, respectively. Due to the large number of buoys in the Northern Hemisphere, these have been validated for each data source with all observations weighted equally. The last column refers to the RMS obtained by validating all test stations in one long time series weighting all daily observations equally. In the Northern Hemisphere the satellite-derived IAT agrees better with *in situ* observations for both surface types than does ERA-Interim. Over the Greenland Ice Sheet the satellite-derived IAT performs better than ERA-Interim, while the northern Alaska site Barrows agrees better with ERA-Interim. Over sea ice ERA-Interim agrees better with in situ observations from the ECMWF data stream and Polarstern. However, the buoys from the ECMWF data stream may have been assimilated into ERA-Interim. The independent *in situ* observations from ACSYS, CRREL, DAMOCLES and FRAMZY are better reproduced by the satellite-derived IAT.

Due to limited observations over the Southern Hemisphere sea ice all buoy data have been included in Figure 3.4.12. Unfortunately, the Southern Hemisphere sea ice observations are limited to Polarstern and those obtained from the ECMWF data stream, where the latter may have been assimilated into ERA-Interim. For land ice, NCEI (NCDC) IAT observations are better reproduced by ERA-Interim, while AMRC are more equally well-reproduced by ERA-Interim and the satellite-derived IAT. In one long time series the satellite-derived IAT is slightly better than ERA-Interim.
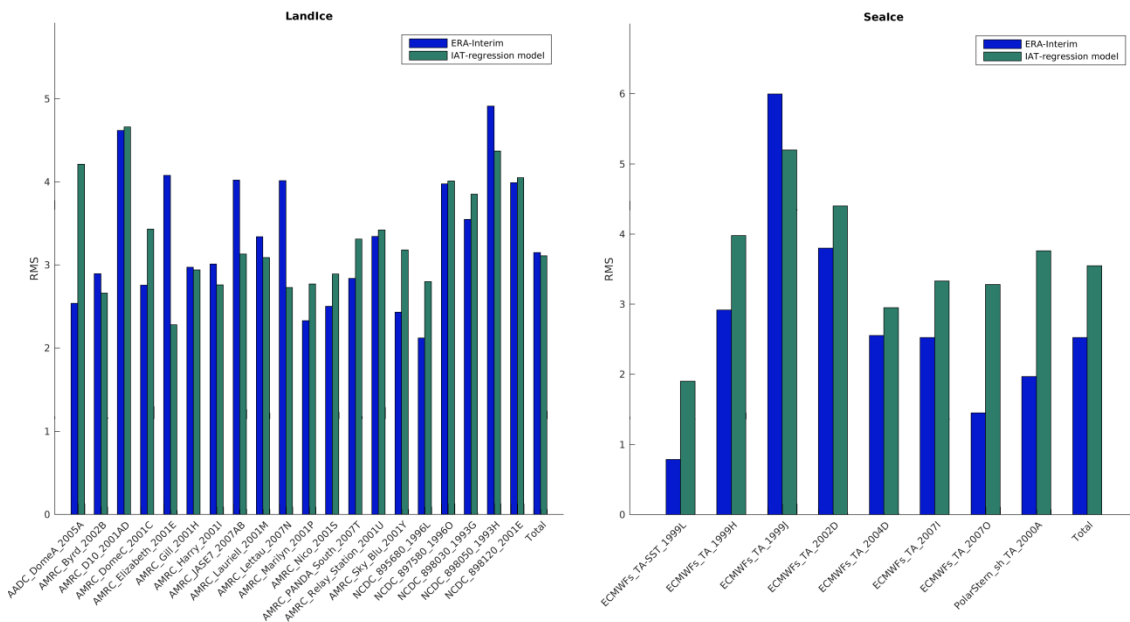
Figure 3.4.12: *Root mean square differences (°C) calculated for the Southern Hemisphere land ice and sea ice, respectively. Blue columns are RMS obtained by comparing in situ with ERA-Interim while green columns are in situ compared with the satellite-derived IAT-data set. The last columns listed as "total" are the RMS obtained by using all test data.*

3.4.7 Conclusions and outlook

The results obtained for the ice covered areas show that there is a large potential for using satellite observed surface temperatures for estimating near surface air temperatures. The satellite derived estimates of IAT validate similarly or better (in some areas) as compared with NWP estimates. In addition, the satellite-derived IAT is independent of NWP models and thus represents an important alternative to the model-based IAT estimates. The regression models presented here all work on satellite observations that are available in near real time and open up opportunities for a near real time estimation of IAT from satellites.

With the sparse number of in situ observations in the Polar Regions, it is not a simple task to gather in situ observations for testing and tuning of the regression models. Large efforts have been put into collecting and quality controlling in situ observations from various data

portals and research projects. Even more work would have resulted in more data, but such efforts are outside the scope of this project.

The work demonstrated that it is crucial for the outcome to have a good reference data set that can be used to understand the conditions and determine the form of the regression model and reliable coefficients. The PROMICE network is an example of a reference data set for the Greenland Ice sheet which was very useful and improved the results. Similar reference data sets in other regions would be useful and would most likely lead to improvements in the IAT estimation.

Future updates (outside EUSTACE) to the IAT based data set could include the addition of satellite and in situ observations up to present. This extension of the data sets would allow for inclusion of radiometric skin surface temperature observations from the sea ice off Qaanaaq gathered by DMI and from the Arctic Sea ice Pack, gathered by Met.no. In addition, communications are ongoing within the EUMETSAT OSI-SAF, to collect observations from the Antarctic Sea ice. These observations would allow for a more detailed training of the regression models and could probably lead to improved IAT estimation in a possible follow on project to EUSTACE.

3.4.8 References

Ahlstrom, A.P. & PROMICE project team (2008). A new programme for monitoring the mass loss of the Greenland ice sheet. Geological Survey of Denmark and Greenland Bulletin 15, 61–64.

Dee, D. P. et al. (2011), The ERA-Interim Reanalysis: Configuration and Performance of the Data Assimilation System, Quarterly Journal of the Royal Meteorological Society, Vol. 137, No. 656, 2011, pp. 553-597. doi:10.1002/qj.828

Dybkjær, G., J. L. Høyer, R. Tonboe, S. M. Olsen (2014). Report on the documentation and description of the new Arctic Ocean dataset combining SST and IST. NACLIM Deliverable D32.28.

Lüpkes, C., Vihma, T., Birnbaum, G., & Wacker, U. (2008). Influence of leads in sea ice on the temperature of the atmospheric boundary layer during polar night. Geophysical Research Letters, 35(3).

Tetzlaff, A., Kaleschke, L., Lüpkes, C., Ament, F., & Vihma, T. (2013). The impact of heterogeneous surface temperatures on the 2-m air temperature over the Arctic Ocean under clear skies in spring. The Cryosphere, 7(1), 153-166.

Vihma, T., & Pirazzini, R. (2005). On the factors controlling the snow surface and 2-m air temperatures over the Arctic sea ice in winter. Boundary-layer meteorology, 117(1), 73-90.

Vihma, T., Jaagus, J., Jakobson, E., & Palo, T. (2008). Meteorological conditions in the Arctic Ocean in spring and summer 2007 as recorded on the drifting ice station Tara. Geophysical Research Letters, 35(18).

3.5 Relationships over Lakes

3.5.1 Introduction

The role of lake surface water temperature (LSWT) measurements from satellite and in situ sources in EUSTACE was originally planned to be directly analogous to the use of land or ice surface temperature (LST or IST) measurements from satellite: i.e., relationships between skin LST/IST and surface air temperature (SAT) are established and used to invert satellite skin observations to SAT estimates (Tmean, Tmax and/or Tmin) applicable to the same day. However, it was understood from prior to the project that LSWT-SAT relationships are somewhat different because of the contrast in the effective thermal inertia of lakes compared to these other surfaces. The effective thermal inertia is the heat capacity per unit area, C_eff, applicable in the relationship between net heat flux per unit area, Q and time-rate of change in surface temperature that occurs in response. Solid surfaces heat and cool over a very shallow depth on short (e.g., daily) timescales because they are limited to molecular thermal conductivity. Liquids mix because of buoyancy and/or mechanical disturbance, transferring heat vertically with an effective conductivity orders of magnitude greater than molecular conduction. This produces a correspondingly large C_eff. LSWTs therefore respond to SAT variations more slowly. The upshot is that radiometric skin temperatures on ice and (many) land surfaces are tightly coupled to (among other factors) SAT, such that a change in SAT (due to the passage of a weather front, for example) influences the skin temperature on a time scale that is short relative to the daily timescale of analysis required in EUSTACE, often in minutes. In contrast, lakes integrate the effect of SAT variations on timescales of many days, as will be elucidated below in the description of our adaptation of a published model of lake-air relationships.

The initial plan in EUSTACE was nonetheless to invert lake temperatures to give an SAT estimate that would apply as a weighted estimate over many days. Use of such an estimate in the context of a daily analysis would be somewhat more complex that use of a daily value, but concepts and definitions were established that could in principle accommodate such data as input to the EUSTACE analysis.

It is now proposed to make a different use of satellite observations of LSWT that we think is more useful to a century-scale analysis, which will be described in outline below. Briefly, it is to use LSWT observations to tune a representation within the EUSTACE analysis of local lake effects on SAT (where local here means over and adjacent to lakes, but also non-negligible at scales of 0.25° latitude/longitude, i.e., not concerning effects that are small compared to the scale of the EUSTACE analysis). The motivations for this modification of approach are:

- to minimise the discontinuity in the degree to which lake effects are present in the SAT analysis prior to and after the advent of the satellite LSWT record, by devising an approach that can be used throughout the analysis period
- to better capture the spatial nature of lake effects on SAT
- because the number of lakes with useful pre-satellite-era LSWT or SAT records, always expected to be small, has turned out to be even smaller than anticipated. Specifically, daily measurements of SAT above lakes are extremely few (a dozen locations for only recent decades, perhaps) compared to the several hundred lakes globally expected to have a non-negligible effect at 0.25° resolution.

The approach to LSWT-SAT relationships now being pursued is presented in the following sections.

3.5.2. Interactions at the air-water interface

The dynamic coupling between a lake and the atmosphere depends on the transfer of momentum, heat and material, at the air-water interface. The magnitudes of these fluxes influence physical processes within lakes which, in turn, have a major impact on the ecosystem. Via local weather, LSWT is strongly influenced by large-scale climatic forcing. The main atmospheric variables that influence lake surface water temperature (LSWT) are surface air temperature (SAT), solar radiation, cloud cover, wind speed, and relative humidity. The thermal interaction between a lake and the atmosphere is complex, as several distinct processes contribute to the heat flux and the resulting change in LSWT (Woolway et

al., 2015). The net heat flux, Hnet, into a lake results from the combination of individual heat flux terms, which can be described by the following equation (shown as the heat flux per unit surface in W m-2):

$$H_{net} = H_s + H_a + H_w + H_e + H_c + H_p + H_i + H_d,$$ (Eq. 3.5.1)

where H_s is the net short-wave radiative heat flux due to solar radiation (considering only the incoming radiation that is actually absorbed), H_a is the net long-wave radiation emitted from the atmosphere toward the lake, H_w is the long-wave radiation emitted from the water, H_e is the latent heat flux (via evaporation/condensation), H_c is the sensible heat flux (via convection), H_p is the heat flux due to precipitation onto the water surface, H_i is the effect of the throughflow of water by inlets and outlets, and H_d is the heat flux exchange at the interface between the epilimnion (the surface layer in a thermally-stratified lake) and hypolimnion (the dense, bottom layer of a thermally-stratified lake). H_p, H_i, and H_d are often neglected when evaluating the influence of lakes on the overlying atmosphere (and vice-versa) and equation (3.5.1) can be simplified to:

$$H_{net} = H_s + H_a + H_w + H_e + H_c,$$ (Eq. 3.5.2)

where at any given time, LSWT tends towards an equilibrium temperature (Teq), defined as the temperature in which net heating/cooling (Hnet) is zero. Thus, lakes are often in a constant state of adjustment with the atmosphere. In equatorial lakes, where the intensity of shortwave radiation and atmospheric conditions both display relatively weak annual signals, lakes stay close to being in thermal equilibrium with the atmosphere. In this case, the individual heat flux terms may be large, but the net heat flux is relatively small. Thus, the SAT-LSWT relationship will be characterized by smaller difference and less variability in the tropics. In contrast, lakes at middle-latitude regions are constantly in a state of adjustment toward thermal equilibrium with the atmosphere. The time scale on which LSWT, for a particular lake, can react to the equilibrating water-air heat fluxes and the progression of the annual cycle in shortwave forcing determines how close to equilibrium the LSWT will be. If the time scale is long (a week or more), the lake may constantly be adjusting to atmospheric conditions and never truly reach equilibrium. If a lake never reaches equilibrium the dynamic coupling between it and the atmosphere is complex, whereas if LSWT is at or near T_eq the relationship between SAT and LSWT can be calculated with relative ease. Thus, for some lakes, the relationship between SAT and LSWT will be one of relatively close coupling, and given available SAT and LSWT data, a relationship can be calculated, allowing the influence of LSWT on SAT, and vice versa, to be determined. One complication for global scale

analyses back to the 19th Century, however, is the sparsity of useful pre-satellite-era LSWT. Thus, a method is firstly needed to estimate LSWT through time.

The influence of LSWT on SAT is not limited to the air immediately above the lake. It has been shown that lakes of sufficient extent tend to moderate SAT anomalies in their surroundings (e.g. Bonan, 1995; Hostetler et al., 1994; Krinner, 2003; Lofgren, 1997; Long et al., 2007). This occurs due to the effective thermal inertia of a lake on daily to monthly timescales being much greater than surrounding land surfaces. Specifically, lakes undergo distinct diel (Woolway et al., 2016) and seasonal (Layden et al., 2015) cycles of surface temperature that differ from the surrounding land, primarily as a result of the large variations in surface roughness, heat capacity, surface albedo, effective heat capacity and energy exchange among different surfaces. Therefore, to capture the influence of lakes (i.e. LSWT) on SAT one needs to not only develop a method that accounts for the relationship between LSWT and the SAT directly overhead, but the method should also consider the spatial influence, thus the region of influence (ROI), of the lake on its surroundings. In the following, we provide information for methods in development for (i) estimating LSWT from readily available SAT data, and (ii) estimating the ROI of a lake on its surrounding SATs.

3.5.3 Estimating LSWT - Modified Kettle model

To estimate LSWT, we modify a previously published LSWT model, selected from a review of possible simple models because it has a minimal number of parameters, focuses directly on the SAT-LSWT relationship, and treats terms not linking SAT and LSWT implicitly. Moreover, the selected model can be readily parameterized for a particular lake from satellite LSWT and in situ or NWP air temperature. Specifically, we modify the commonly used Kettle model (Kettle et al., 2004). In its original form, the model estimates LSWT from a lagged and smoothed time series of SAT, and can be described as approximating the slow integrated response of LSWT to meteorological forcing, of which SAT is the primary driver. The model uses SAT observations to estimate LSWT by assuming that LSWT, $T_{w,t}$, is controlled predominantly by a combination of smoothed SAT, $f(T_a,t)$, and clear-sky radiation, S. Specifically, $T_{w,t}$ can be modelled as a linear combination of these variables as:

$$T_{w,t} = a + bf\left(T_{a,t}\right) + cS,$$
(Eq. 3.5.3)

where

$$f\left(T_{a,t}\right) = \alpha T_{a,t} + (1 - \alpha)f\left(T_{a,t-\Delta t}\right),$$
(Eq. 3.5.4)

is an exponential smoothing function with the smoothing parameter, α, in the interval [0, 1]. As described by Kettle et al. (2004), a, b, c, and α are determined using an iterative process that gives the best fit to the observed LSWT in terms of a mean absolute difference (MAD) statistic. In EUSTACE, we modify this approach to consider the known LSWT climatology (from satellite observations) of a given lake. Specifically, using the notation $X = X' + \bar{X}$, where X' is used to denote the anomaly and $\bar{X}$ is used to denote the climatology, we can re-formulate the Kettle model as:

$$T_w = T_w{}' + \overline{T_w} = a + bf(T_a{}' + \overline{T_a}) + cS, \qquad \text{(Eq. 3.5.5)}$$

where the LSWT anomaly, $T_w{}'$, can be expressed as:

$$T_w{}' = bf(T_a{}'), \qquad \text{(Eq. 3.5.6)}$$

and the LSWT climatology, $\overline{T_w}$, as:

$$\overline{T_w} = a + bf(\overline{T_a}) + cS, \qquad \text{(Eq. 3.5.7)}$$

which gives:

$$T_w = \overbrace{bf(T_a{}')}^{\text{Anomaly}} + \overbrace{a + bf(\overline{T_a}) + cS}^{\text{Climatology}}. \qquad \text{(Eq. 3.5.8)}$$

However, as we know $\overline{T_w}$ from observations, primarily from satellite analyses (MacCallum and Merchant, 2012), equation 3.5.8 can be simplified further to:

$$T_w = bf(T_a{}') + \overline{T_w}, \qquad \text{(Eq. 3.5.9)}$$

which removes the *cS* term and relies instead on the observed $\overline{T_w}$, which imposes a more realistic form of the annual radiative cycle, implicitly. Therefore, our equation simplifies to only one unknown, $f(T_a{}')$. The new formulation, similar to Kettle et al. (2004), is then solved iteratively to obtain *b* and *a*, which are determined according to the modelled T_w that provides the best fit with the observed LSWT in terms of MAD.

3.5.4 Example for Lake Superior

To demonstrate the usefulness of the modified Kettle model, we use one of the most complete LSWT and SAT datasets available from anywhere in the world. We concentrate on the data available for Lake Superior (North America), one of the world's largest and most well-studied lakes. The LSWT and SAT data used from Lake Superior were collected from

the National Data Buoy Centre (NDBC), which maintains three surface buoys in Western, Central, and Eastern Lake Superior. These monitoring stations (45001, 45004, 45006) record LSWT and SAT with a reported accuracy of better than 1°C and a resolution of 0.1°C. They are often deployed in March or April, before the start of the thermally stratified season, and are removed in October or November, before the start of ice formation. An uninterrupted LSWT series is also available from the ATSR (Along Track Scanning Radiometer) Reprocessing for Climate: Lake Surface Water Temperature and Ice Cover (MacCallum and Merchant, 2012) dataset (ARC-Lake v2), which contains observations from August 1991 to December 2009, derived from the ATSR series, which consists of ATSR-1, ATSR-2, and Advanced(A)ATSR at a spatial resolution of 0.05°. For validation of ARC-Lake data with in situ LSWT temperature we extract LSWT for the pixel (0.05°) nearest to each buoy. These measurements were then compared to daily average NDBC buoy temperatures for all days when they were in operation. The three monitoring stations on Lake Superior had a MAD of 057, 0.56, and 0.62 for buoys 45001, 45004, and 45006, respectively.

As shown in Figure 3.5.1 SAT and LSWT in Lake Superior vary considerably throughout the seasonal cycle with substantial intra-annual variability. Comparison of daily mean SAT and LSWT reveals them to have essentially the same temporal structure at the seasonal level, but varying considerably at shorter timescales. The seasonal cycle of LSWT lags that of SAT during spring, but remains warmer during fall and winter. We see clearly how LSWT lags SAT in the early part of the year and LSWT remains warmer for longer in the latter stages of the seasonal cycle (Fig. 3.5.1b).
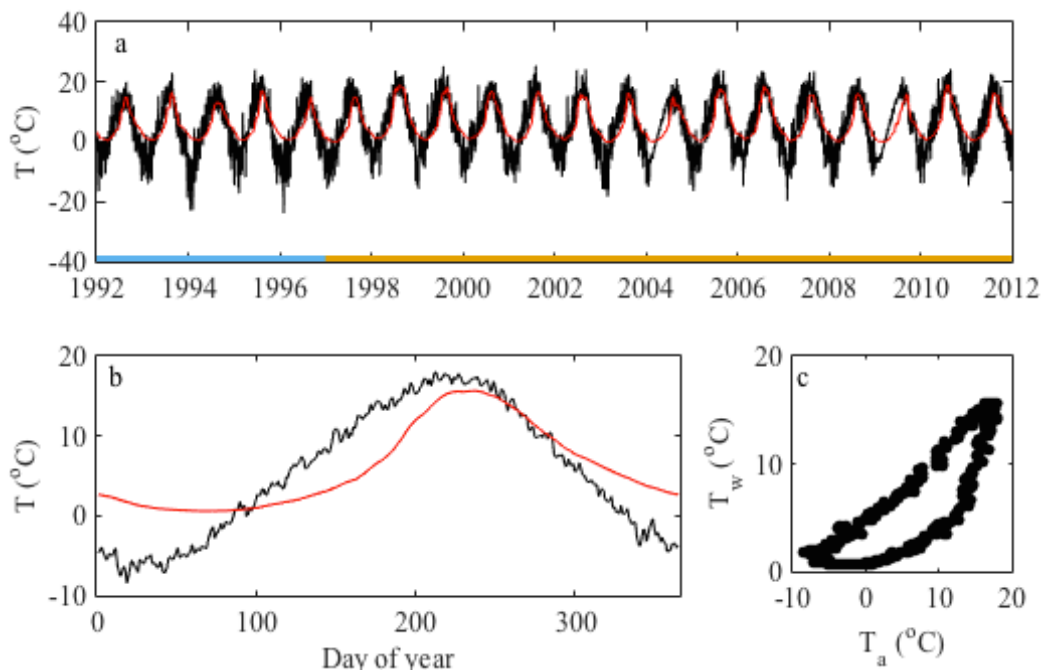
Figure 3.5.1. *Time series of (a) daily-averaged lake mean surface water temperature (LSWT, °C; red) and surface air temperature (SAT, °C; black); (b) comparison of LSWT (red) and SAT (black) climatology; and (c) hysteresis cycle of LSWT and SAT climatology. The calibration period for the method presented in this investigation is shown by the light blue section in (a) and the validation period is shown in orange.*

The modified Kettle model simulates LSWT in Lake Superior with success (Figure 3.5.2). Specifically, we see that the model captures the inter-annual variability in the observed LSWTs and follows a clear seasonal cycle, never deviating much from the observed temperatures (MAD = 0.6ºC). The modified Kettle model captures this temporal variation including the lagged response of LSWT to SAT.
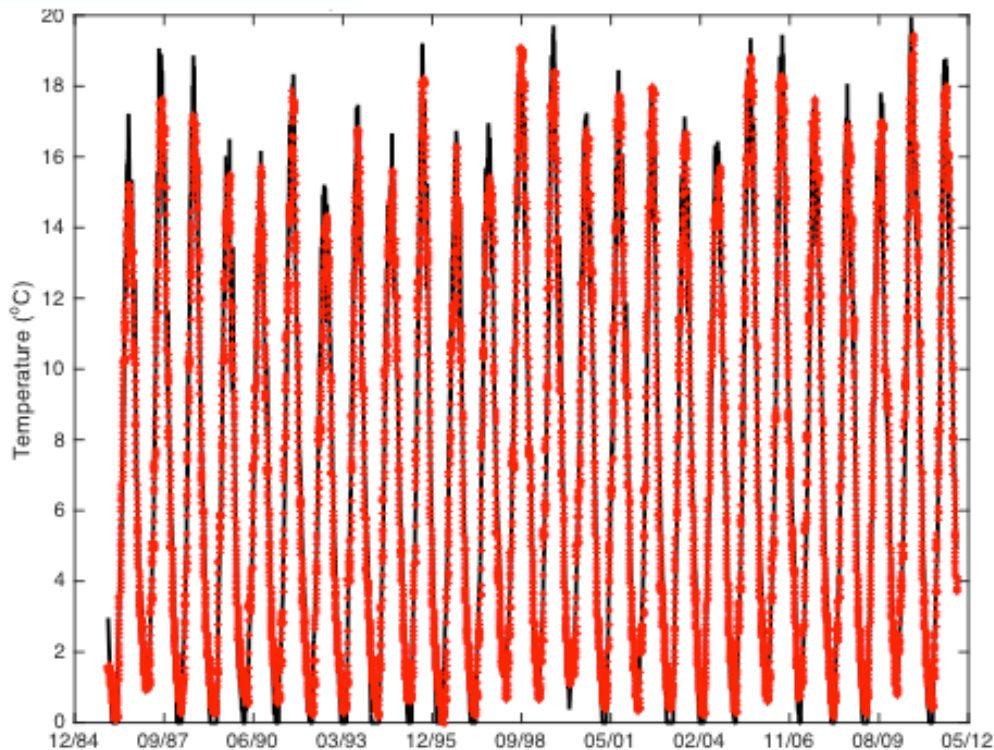
Figure 3.5.2. *Example of simulated (red) LSWTs (°C) from the modified Kettle model with the observed LSWTs from NDBC (black).*

As LSWTs vary spatially within a lake, one would also expect that the model parameters in the modified Kettle model would vary across the lake surface. To evaluate the fine-scale variations in each model parameter, we simulate (using the SAT from NDBC) LSWT at each 0.05º grid box (as available in the ARC-Lake data). We compare the spatial variations in each model parameter and the MAD with lake bathymetry data, available from the Global Lake Database (GLDB, v2), for which data were either digitised from different topographic maps or extracted from ETOPO1, a 1 arc-minute relief model of Earth's surfaces (see Choulga et al., 2014). We find considerable spatial variations in the model parameters and in the model performance (i.e. MAD). However, our results demonstrate that the modified Kettle model works considerably well to generate a LSWT time series in Lake Superior and that none of the model parameters vary significantly with depth, and MADs were < 1ºC throughout. By using the model coefficients from the lake mean simulations, and modelling LSWTs spatially within Lake Superior, where the only differences among model runs were

the LSWT climatology, we find that the MAD in each 0.05º box was still below 1ºC. This implies an influence of the LSWT climatology for determining the accuracy of the Kettle model. One limitation of the above analysis is that one would ideally use spatially resolved SATs at the same spatial resolution as the LSWTs. However, these are not available (for example, SAT from reanalysis are often available at a spatial resolution of 0.75º) and thus SAT from the lake centre were used.

3.5.5 Example for the African Great Lakes region

To investigate further the usefulness of the modified Kettle model for simulating LSWTs, we simulate LSWTs for the African Great Lakes (AGL) region. For the AGL, we use two temperature datasets: SAT and LSWT. Satellite LSWT observations are available from the ARC-Lake dataset. SAT observations in the tropics are limited, especially at the lake surface. To our knowledge, no long-term datasets of SAT above any of the AGL are available, and in this study we are limited to only a few SAT observations available from a very few meteorological stations. Thus, we need to estimate SAT above the lake in order to derive an SAT-LSWT relationship. To determine SAT above and near the lakes, we use a regional climate model (RCM). Specifically, we use the Consortium for Small-Scale Modeling model in climate mode (COSMO-CLM) coupled to the Freshwater Lake model (FLake) and Community Land Model (CLM), to dynamically downscale a simulation from the African Coordinated Regional Downscaling Experiment (CORDEX-Africa) to 7-km grid spacing. The models involved in this study are described in detail by Thiery et al., (2015). SAT estimates from the RCM agree well with observed SAT from nearby meteorological stations (available from local airports through personal data contacts [not publically available] and from the GHCN-D), with a mean difference of 0.4K and a RMSE of 0.9K. However, unlike the meteorological station data, the model data is generated at a temporal resolution of 3 hours and provides an uninterrupted temperature series thus can be used to estimate LSWTs from the modified Kettle model.

Prior to running the modified Kettle model, we first explore the LSWT-SAT relationships for one year (2002) at a 3-hour temporal resolution. We demonstrate that the ratio of SAT to LSWT is very close to unity in these lakes, with only minor deviations at high temporal resolution (Figure 3.5.3).
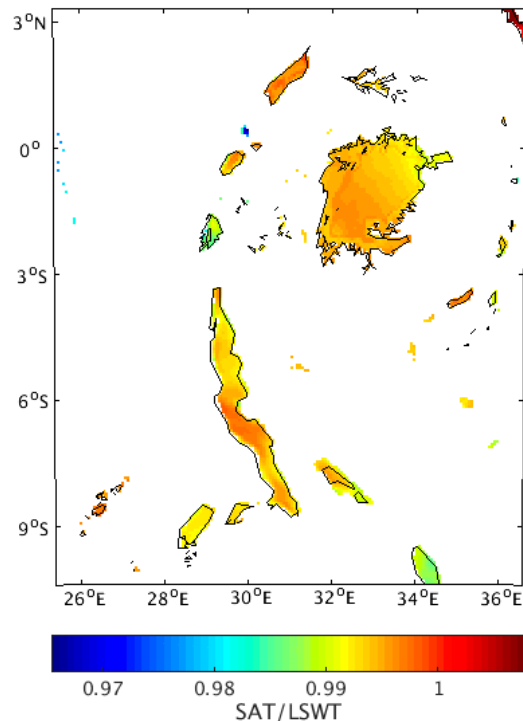
Figure 3.5.3: Annually averaged ratio of air (SAT) to water (LSWT) temperatures in the African Great Lakes region.

Minor within-lake differences in the ratio of SAT to LSWT are observed, where in particular the shallowest regions in Lake Tanganyika demonstrate closer LSWT to SAT relationships. These spatial patterns are driven by the spatial variations in the depth of the upper mixed layer (Figure. 3.5.4), where shallower mixing depths result in closer associations between LSWT and SAT. This is consistent with the high co-variability known to exist between LSWT and SAT in shallow lakes (and hence shallow mixing depths). In deep lakes (or deeper mixed layers), in contrast, variations of LSWT driven by variations of SAT are highly dampened (Toffolon et al., 2014), due to the greater thermal inertia (i.e. the increased volume of water that is affected by surface heating). Similar patterns are observed throughout the seasonal cycle, with differences in the depth of the upper mixed layer being highly correlated with the SAT/LSWT relationship.
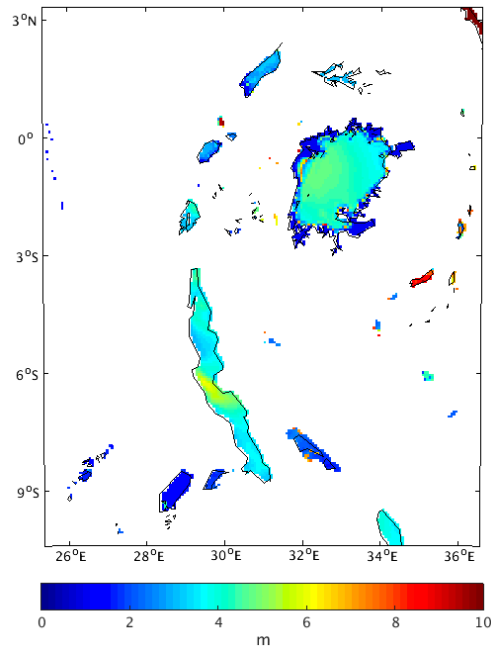
Figure 3.5.4: *Annually averaged depth (m) of the upper mixed layer (epilimnion) in the African Great Lakes.*

For tropical lakes, we conclude that a direct linear relationship between SAT and LSWT exists, so LSWT can be used directly to estimate SAT, using a constant close to unity (i.e. 1:1) for daily mean temperatures. The same is valid for daily minimum and daily maximum temperatures as shown in Figure 3.5.5. One of the main issues is the lack of in situ meteorological data above the water surface in these lakes; the only data available are from nearby meteorological stations (i.e. not above the lake surface).
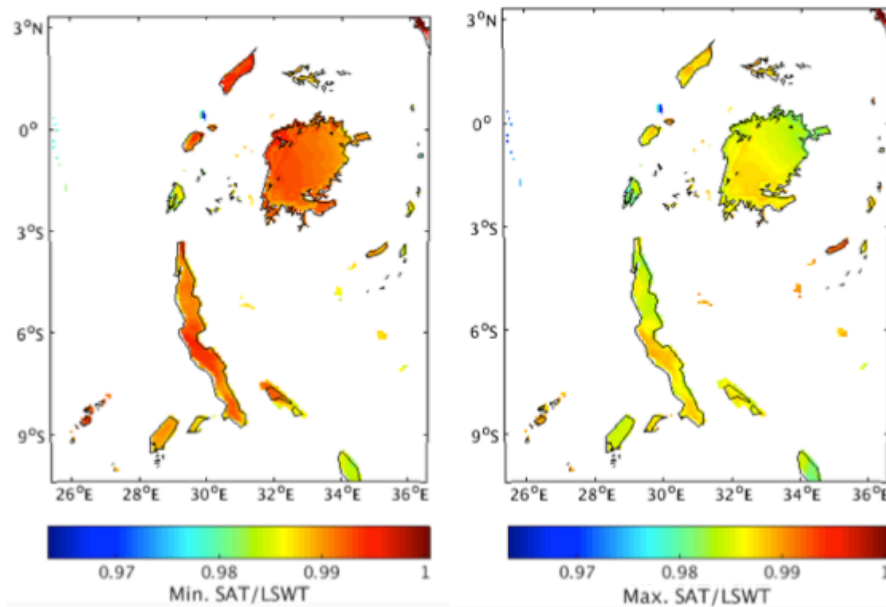
Figure 3.5.5: *(Left) Annually averaged ratio of the daily minimum air (SAT) to water (LSWT) temperatures in the African Great Lakes region. (Right) Annually averaged ratio of the daily maximum air (SAT) to water (LSWT) temperatures in the African Great Lakes region.*

To determine long-term LSWTs at daily resolution, we use the modified Kettle model. Model outputs demonstrate strong agreement between simulated and observed temperatures in Lakes Victoria, Tanganyika, Albert, and Kivu, with an average MAD <0.5ºC. Running the modified Kettle model for 263 lakes from around the world, driven with ERA-Interim data, demonstrates that the majority of lakes had a MAD of less than 1ºC (Figure 3.5.6) when compared to ARC-Lake.
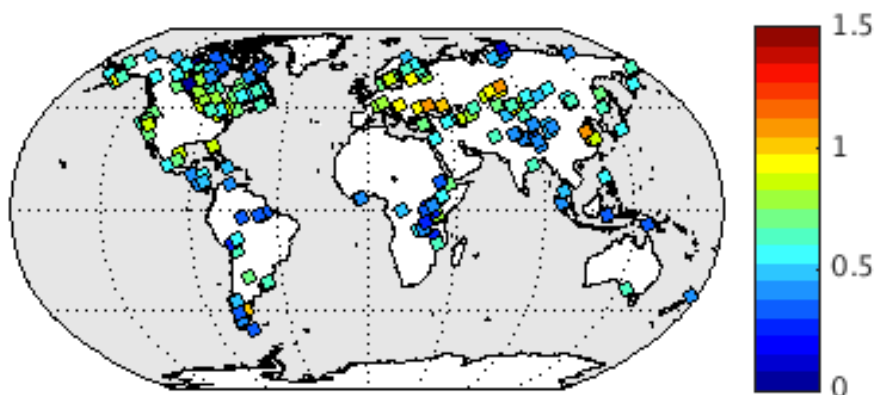
Figure 3.5.6: *Error statistics for the modified Kettle model when used for a global (n = 263) dataset. Colours signify MAD (°C).*

3.5.6 Spatial effects of lakes on surface air temperatures

As mentioned in section 3.5.1, lakes also influence the climate of their surroundings, in particular in regions where they are abundant. Mediated by the fluxes of energy, moisture, and momentum, they significantly alter the surface energy and water balance in a particular region and therefore its climate. Here, we demonstrate the influence of lakes on the climate and introduce a possible method for their inclusion in the EUSTACE analysis by comparing model simulations for the African Great Lakes region with and without the presence of lakes. That is, the results of the RCM simulations described in section 3.5.3 are compared to a no-lake simulation. The with-lake and without-lake simulations are identical, but for the no-lake simulations each lake pixel is replaced by a representative land pixel selected from all land pixels within a distance of 50km. This approach makes it possible to quantify the impact of lakes, which with their different surface roughness etc. will result in variable surface heat exchange compared to land, on the regional climate, in particular near-surface temperature.

The climatological difference between the two simulations is seen to be up to 2°C, and the African Great Lakes tend to induce a cooling effect, on average (calculated as the annual average from 3-hour simulations). Only one meteorological station, with available observational data, appears to be strongly influenced climatologically (effect > than ~0.2°C) by the presence of these large lakes (Figure 3.5.7). The spatial structures of climatological influence are related to topography, and are at scales not resolved by the station density. Thus, neglecting to include lake effects when infilling between SAT stations can result in a substantial bias in SAT estimates.
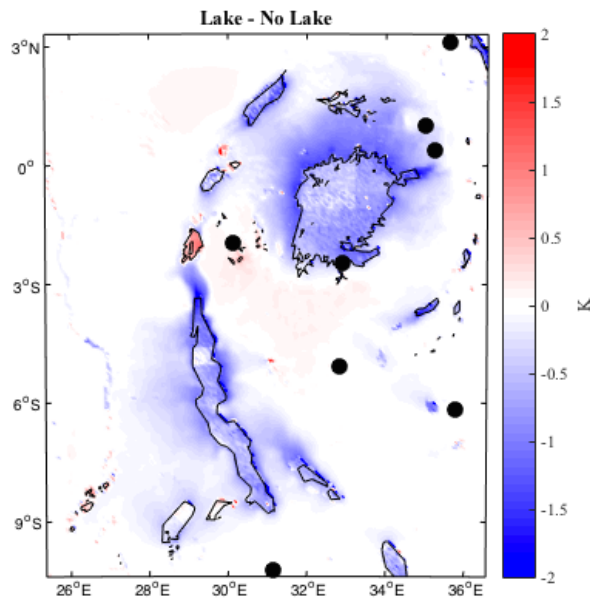
Figure 3.5.7*: Mean difference in SAT (°C) from running identical model (and boundary forcing of synoptic situation) with no lakes (replacing them with appropriate land surface, as described in the text). Black markers show the location of the GHCN-D stations with available data.*

For EUSTACE, the key question is how the lake modifies the dynamics over time of the daily min, max, and mean SAT in its vicinity. We have explored this by deriving the leading EOFs (Empirical Orthogonal Functions) for Lake Victoria (and nearby Lake Albert) of the daily lake-no-lake model differences after removing monthly-scale variability by Gaussian smoothing, Figure 3.5.8. These results illustrate the 'average' influence of the lake on its surroundings, neglecting the day-to-day variability which arises as a result of differences in wind speed/direction and cloud cover etc.
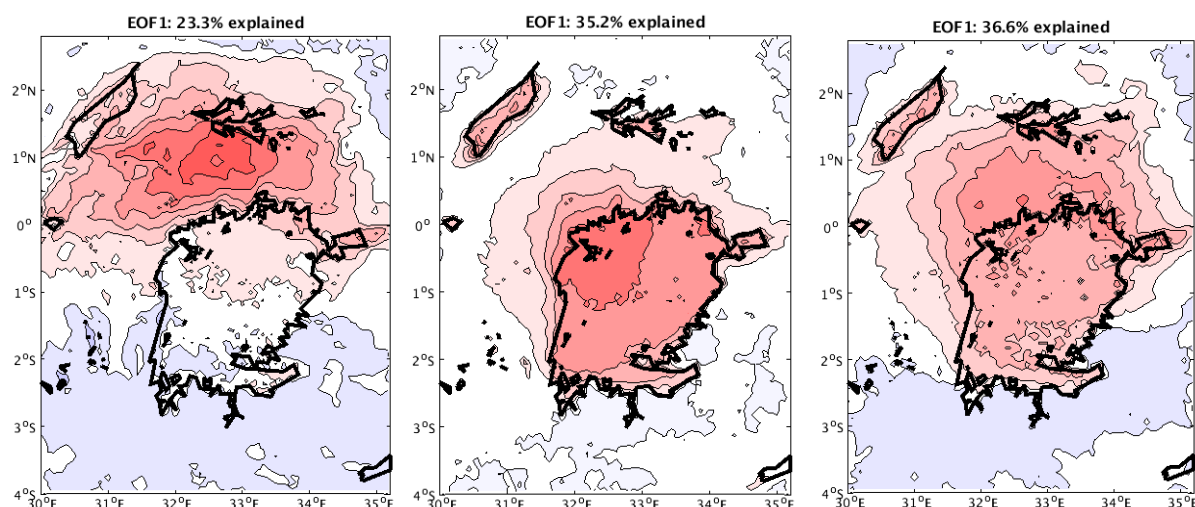
Figure 3.5.8:*. EOF1 of daily lake-no-lake modelled differences for the Victoria region (for the same period of synoptic weather conditions). Left: daily min SAT. Middle: daily max SAT; Right: daily mean SAT. Sign attributed to pattern is arbitrary. The mean value has been calculated as a 24-hour mean rather than as the mid-value between max and min.*

The influence of lakes on the surrounding SAT is seen clearly for Lake Victoria from the EOFs (Figure 3.5.8). Next, we calculate the region of influence (ROI) of Lake Victoria on its surrounding area and generate a test dataset, which could be used within the infilling method of EUSTACE. To determine the ROI, we calculate the region in which the lakes 'significantly' affect the SAT observations. We define a significant ROI as the region in which the presence of lakes increases or decreases the surrounding SAT by 0.5ºC (Figure 3.5.9), calculated after converting the EOFs to have units of temperature.

The ROI illustrates that the SAT directly above and to the north of Lake Victoria is significantly influenced by the lake presence. Specifically, the region of maximum influence is to the north of the lake, clearly a topographic effect and driven by a lake-breeze system. To determine the magnitude of influence of the lake, we use the following relationship:

$$\delta T_a = c(T'_w - T'_a) \qquad\qquad\qquad\qquad\qquad \text{(Eq. 3.5.10)}$$

where δT_a represents the lake-effect modification and T'_w and T'_a are the anomalies in water and air temperature, respectively, averaged over the lake surface. T'_w can be calculated from the ARC-Lake dataset or, where no data is available, from the modified Kettle model, where

temperatures are averaged across the lake surface, thus representing the lake mean temperature. The SAT anomaly, T_a', can be defined as the sum of the SAT anomaly that would be obtained in the absence of any lake effects, t_n^a, and the lake effect modification of SAT, δT_a:

$$T_a' = t_n^a + \delta T_a \qquad \text{(Eq 3.5.11)}$$

which we can parameterise as:

$$T_a' = t_n^a + c(T_w' - T_a') \qquad \text{(Eq 3.5.12)}$$

from which it follows that

$$T_a' = \frac{cT_w' + t_n^a}{1+c}. \qquad \text{(Eq 3.5.13)}$$

In regions where the lake effect is negligible $T_a' = t_n^a$, and the measured (i.e. from nearby SAT data) values, t_a^m, are used directly as inputs to this analysis (c = 0). This also applies to periods when lakes are frozen, which will not be the case for Lake Victoria, but will occur in winter at higher latitudes. However, when the lake presence has an influence on T_a', as shown by the EOFs, the parameter c, must be calculated. We calculate 'c', which is a cell-specific constant, by firstly calculating the anomalies in SAT for with and without lake model simulations. For cells that are influenced significantly by the presence of lakes, we calculate, by regression, the value of c in $c(t_w - t_a) = t_a - t_n^a$, where $t_a - t_n^a$ can be summarized by the time series of the first principal component of the EOF. The 'c' parameter is then multiplied to each cell within the ROI, which have been normalized from 0 - 1, and used in Equation 3.5.13 (Figure. 3.5.9).
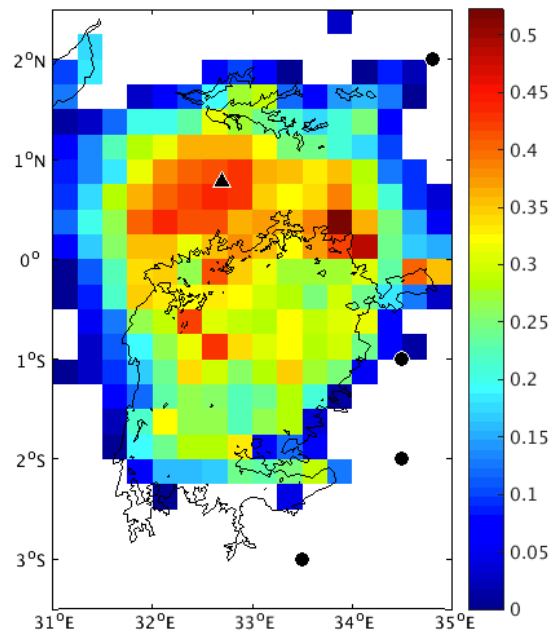
Figure 3.5.9: *Region of Influence of Lake Victoria on the surrounding surface air temperature (SAT). Red signifies large influence and blue signifies small influence. Black markers illustrate locations where SAT values are extracted from the model for the EUSTACE test case. The value within each cell will be used in Equation 3.5.13 to calculate the air temperature anomaly with the presence of lakes.*

3.5.7 Conclusion and outlook

The influence of lakes on the surrounding SAT can be substantial, and in some instances be in excess of 2°C. Thus, in some regions, in particular where lakes are abundant, their influence on the surrounding climate needs to be considered. For EUSTACE, the key question is how the lake modifies the dynamics over time of the daily min, max, and mean SAT in its vicinity. (In the above analysis, we demonstrate the ROI on daily mean SATs, but similar maps are generated, and used in the analysis, for the minimum and maximum daily SATs. Thus, three different ROIs are calculated for evaluating the lake effect on regional SATs.) We have explored this by deriving the leading EOFs (Empirical Orthogonal Function) of the daily lake-no-lake model differences from a regional climate model output. Our results demonstrates that the dominant EOFs have spatial patterns that are clearly strongly influenced by the local topography, as observed by the spatial patterns of the ROI (for

example, the region south of Lake Victoria is not influenced by lake anomalies as a result of the presence of high ground limiting the lake influence). Previously, the influence of lakes on the surrounding SAT has been neglected in observational studies, specifically when observational data are not available and an infilling method is used. By following the approach presented here we demonstrate how the influence of lakes on SATs can be included and potentially allow for a greater accuracy of SAT estimates when meteorological observations are scarce. By running a RCM with and without the inclusion of lakes we demonstrate that a region of influence can be computed and then could be used within an infilling approach throughout the analysis period to better capture the spatial nature of lake effects on SAT. Moreover, when LSWTs are not available, we demonstrate a method for estimating LSWT from nearby air temperature observations. Specifically, we show that surface air temperatures and the LSWT climatology can be used to generate accurate (MAD < 1°C) time series of LSWTs throughout the world. The method presented can be used to determine LSWT back to the 19th Century. The next goal will be to generate a global dataset of lake ROIs and apply this approach (similar to the African Great Lakes example) to all lakes which 'significantly' influence the climate of their surroundings from around the world.

3.5.8 References

Choulga, M., E. Kourzeneva, E. Zakharova, and A. Doganovsky (2014), Estimation of the mean depth of boreal lakes for use in numerical weather prediction and climate modelling, *Tellus A*, 66, 21295.

Kettle, H., et al. 2004. Empirical modeling of summer lake surface temperatures in southwest Greenland. Limnol. Oceanogr. 49, 271-282.

Layden, A., S. N. MacCallum., and C. J. Merchant (2016), Determining lake surface water temperatures worldwide using a tuned one-dimensional lake model (Flake, v1), *Geosci. Model Dev*. 9, 2167-2189.

MacCallum, S. N., and C. J. Merchant (2012), Surface water temperature observations of large lakes by optimal estimation, *Can. J. Remote Sens.* 38, 25–44.

Thiery, W. et al. (2015). The impact of the African Great Lakes on the Rgional Climate. Journal of Climate 28, 4061-4085.

Toffolon, M., et al. (2014), Prediction of surface temperature in lakes with different morphology using air temperature, *Limnol. Oceanogr.* 59, 2185-2202.

Woolway, R.I., et al. (2015). Automated calculation of surface energy fluxes with high-frequency lake buoy data. Env. Mod. Soft. 70, 191–198.

Woolway RI, et al (2016) Diel surface temperature range scales with lake size. PLoS ONE 11(3):e0152466. doi:10.1371/journal.pone.0152466

3.6 Application of a physically-motivated model for all surfaces

3.6.1 Introduction

The methods used for predicting surface air temperature from satellite observations in EUSTACE have primarily been statistical, whereby relationships over different surfaces have been defined using empirical data and applied to satellite observations at local-to-global scales (sections 3.2-3.4). This is very much in line with what has been documented elsewhere in the published literature (Benali et al., 2012; Chen et al., 2014; Good, 2015; Kilibarda et al., 2014; Janatian et al., 2016; Lüpkes et al., 2008; Oyler et al., 2015; Parmentier et al., 2015; Tetzlaff et al. 2013; Vihma and Pirazzini, 2005; Vihma et al., 2008 and Zhang et al., 2011), and produces acceptable results with quantifiable uncertainties (sections 3.2-3.4). This section summarises trials conducted using a physical model for estimating satellite-derived air temperature, an alternative approach that has received only very limited attention in the published literature (e.g. Sun et al. (2005)).

An overview of the physical model used and the sensitivity of its surface air temperature predictions to various input data is presented in section 3.6.2. In section 3.6.3, the ability of the model to predict instantaneous, sub-daily surface air temperature at a station in Alaska is assessed using only in situ observations to prescribe the required inputs to the model. This gives an approximate upper limit (or 'best-case scenario') of the expected performance of the physical model under the test conditions. The performance of the model in predicting instantaneous 6-hourly surface air temperature over Europe and Africa using satellite LST data is assessed in section 3.6.4. This second experiment provides insight into the feasibility of applying the model on regional to global scales in order to provide satellite-derived air temperature estimates to use in an analysis such as that provided by the EUSTACE project.

3.6.2 Overview of the Physical Model

The physical model adopted here has been built using modules from the JULES (Joint UK Land Environment Simulator) model, which is a community land surface model based on the

Met Office Surface Exchange Scheme (MOSES) (Best et al., 2011; Clark et al., 2011). In JULES, surface air temperature (T2m) is calculated from the skin temperature (*Tskin*), and certain properties of the land surface and lower atmosphere. The JULES modules (version date 24 January 2012) that perform this core task have been aggregated here to produce a stand-alone model that estimates T2m from prescribed atmospheric and surface data:

*T2m =(Tskin - g/cp*1.5 + R * (T1 - Tskin + g/cp *(z1+z0m-z0h)))*   *(Eq. 3.6.1)*

Where *g* is the acceleration due to gravity (9.81 ms^{-1}), *cp* is the specific heat capacity of dry air (1.005 kJ/kg.K), and *T1* is the temperature of the lowest atmospheric layer, which is at height *z1*. The parameter *z0m* is the roughness length for momentum, while *z0h* is the roughness length for heat and moisture. *R* is the ratio of the Monin-Obukov stability functions of observation height (i.e. screen level) over the lowest atmospheric level. Calculating this ratio is the primary computational function of the physical model, which includes calculation of surface transfer coefficients (e.g. heat and moisture, surface scaling velocity, Monin-Obukhov length).

The input parameters required by the physical model are:

- *Tskin* (skin temperature)
- *T1* (temperature of the lowest atmospheric layer available)
- *Z1* (height of *T1*; this varies, depending on the atmospheric temperature data available)
- *z0m* (roughness length for momentum)
- Wind speed at some height above the surface (*W*)
- Height of *W*
- Boundary layer height

The parameter *z0h* (roughness length for heat and moisture) is assumed to be *z0m*/10 (*J. Edwards, Met Office, Personal Communication*). Expected values for *z0m* are shown in Table 3.6.1.

A key consideration in the use of a physical model such as this to estimate satellite-derived air temperatures is the prescription of these input parameters. Information on these variables has to be obtained from somewhere and any source of such data is likely to contain errors, so it is important to understand the sensitivity of the predicted air temperature to those parameters, in order to determine whether or not the method can be reliably used. Figures

3.6.1 and 3.6.2 show the sensitivity of the predicted *T2m* to each of the (non-height) input parameters, keeping all other parameters fixed.

The results show that in these trial cases, there is virtually no sensitivity to the boundary layer height and only a low sensitivity to *z0m* (up to ~1 °C). For *T1>Tskin (Figure 3.6.1)*, i.e. in stable conditions, there is moderate (a few °C) sensitivity to changes in wind speed at low wind speeds, but only a low sensitivity (up to ~ 1 °C) at higher wind speeds, and where *T1<Tskin (Figure 3.6.2)*. In contrast, the sensitivity to *Tskin* for *Tskin<T1*, and to *T1* at all temperatures, is almost one-to-one. For *Tskin>T1*, in unstable conditions, the sensitivity to *Tskin* is markedly lower.

Table 3.6.1: *The Davenport-Wieringa roughness-length classification (Stull, 2015, page 700). The Land Cover CCI classes mapped to each value of z0m used in this study are also shown.*

Class	z0m (m)	Description	Land Cover CCI classes
Sea	0.0002	sea, paved areas, snow-covered flat plain, tide flat, smooth desert	Water bodies (210)
Smooth	0.005	beaches, pack ice, morass, snow-covered fields	Permenant snow and ice (220)
Open	0.03	grass prairie or farm fields, tundra, airports, heather	Grassland (130), lichens and mosses (140), sparse vegetation (150), bare areas (200)
Roughly open	0.1	cultivated area with low crops and occasional obstacles (single bushes)	Croplands (10-30)
Rough	0.25	High crops, crops of varied height, scattered obstacles such as trees or hedgerows, vineyards	Mosaic natural vegetation (40), flooded shrub/herbaceous cover (180)
Very rough	0.5	mixed farm fields and forest clumps, orchards, scattered buildings	Mosaic tree, shrub, herbaceous cover (100, 110), shrubland (120), Flooded tree cover (160, 170)
Closed spaces	1.0	regular coverage with large size obstacles with open spaces roughly equal to obstacle heights, suburban houses, villages, mature forests	Tree cover classes (50-90)
Chaotic	≥2	Centres of large towns and cities, irregular forests with scattered clearings	Urban areas (190)

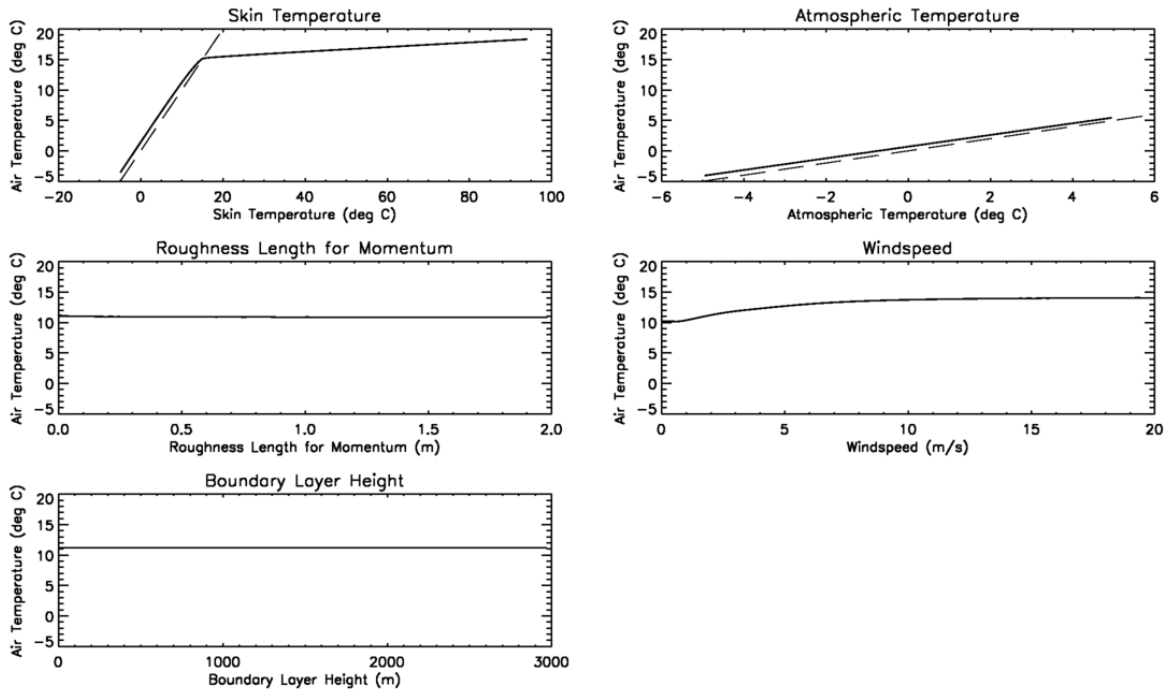**Figure 3.6.1:** *Sensitivity of T2m predicted by the physical model to perturbations in skin temperature (Tskin), the temperature of the lowest atmospheric model level (T1), roughness length for momentum (z0m), wind speed of the lowest atmospheric level (W), and the boundary layer height (BLH). In each case, all parameters are fixed with only the parameter of interest allowed to vary. Fixed parameters: Tskin = 10 °C, T1=15 °C, z0m = 3E-4 m, W=2 ms⁻¹, BLH=1000m. The dashed lines represent the one-to-one line.*
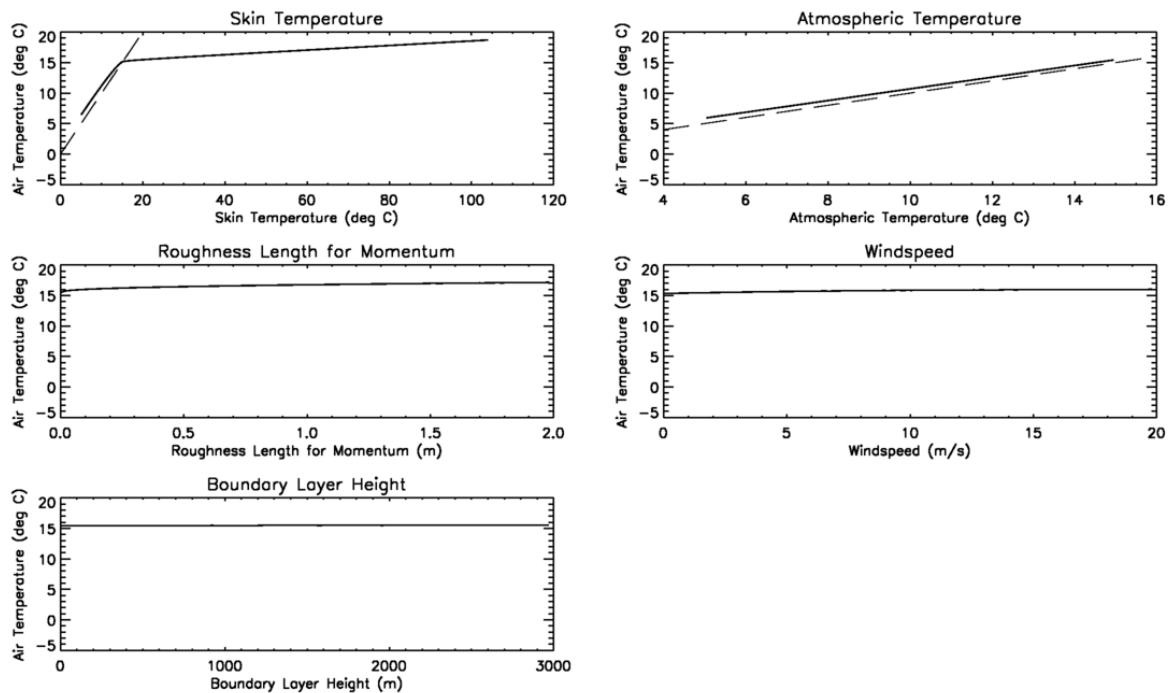
Figure 3.6.2: *As for Figure 3.6.1 but for fixed parameters: Tskin = 20 °C, T1=5 °C, z0m = 3E-4 m, W=2 ms^{-1}, BLH=1000m.*

3.6.4 Test application of the physical model in the Arctic using ARM in situ observations

Application of the model in Arctic conditions is challenged by the very limited amount of in situ observations for validation. In this section, we focus on the NSA Barrow Facility located on the northernmost point in Alaska, USA. This station is unique for its thorough and relatively consistent measuring program, and thus the seasonality, daily cycle and effects of extreme Arctic conditions can be studied at high temporal resolution. In this section, the model is used to calculate full resolution (1-minute) surface air temperature (*T2m*) based on observations of skin temperature (*Tskin*), 20 m temperature (*T1*) and 20 m wind speed (*W*). The predicted *T2m* is validated using in situ *T2m* observations. The analysis is performed on available data from October 2003 to March 2015, for periods with snow covered ground (generally mid-September to start-June).

The Tskin observations are surface infra-red temperatures measured by ground-based radiation pyrometers as a part of the ARM climate research facility in Barrow, available from

https://www.arm.gov/instruments/irt. The 2 m and 20 m air temperature and 20 m wind speed are measured from a meteorological tower at the facility, and data are also available for 10 m and 40 m height (Atmospheric Radiation Measurement (ARM) Climate Research Facility, 2003). All are 60-second averages of measurements made every 1-2 seconds. Finally, daily surface conditions based on in situ measurements of albedo are used to filter snow-only days, to make the results representative for Arctic snow and ice surfaces. *z0m* is assumed to have a fixed value of 0.0005 m, in between the "Sea" and "Smooth" classes of the Land Cover Climate Change Initiative (CCI) data set (Table 3.6.1) (Bontemps et al., 2012, Poulter et al., 2015). A fixed boundary layer height of 1000 m is assumed for all *T2m* calculations. The impact of this assumption is tested below (Figure 3.6.4).

Figure 3.6.3 and Table 3.6.2 shows histograms and statistics of the modelled *T2m* minus the observed *T2m* for the entire dataset 2003-2015, with the observations split into three categories: unstable conditions (*Tskin > T1*), stable conditions (*T1 - 5°C < Tskin ≤ T1*) and very stable conditions (here defined as *Tskin ≤ T1 - 5°C*). Overall, the model performance shows a small cool bias of -0.2°C and a standard deviation of 1.1°C, with a long cold tail, but this is made up of these distinct populations arising from the three stability conditions. The unstable conditions occur in 35% of the samples, and here there is a small warm bias of 0.3°C, and a narrow standard deviation of 0.3°C, with a small warm tail. For stable conditions (54% of all cases), the model has zero bias, but a slightly larger spread (standard deviation of 0.6°C). During the very stable conditions (11% of all cases), the model has less skill, with a cool bias of -2.3°C and root-mean square differences of 3.1°C. This category thus contains the most data with elevated model errors, but still shows a correlation of 0.96 between model and observations.
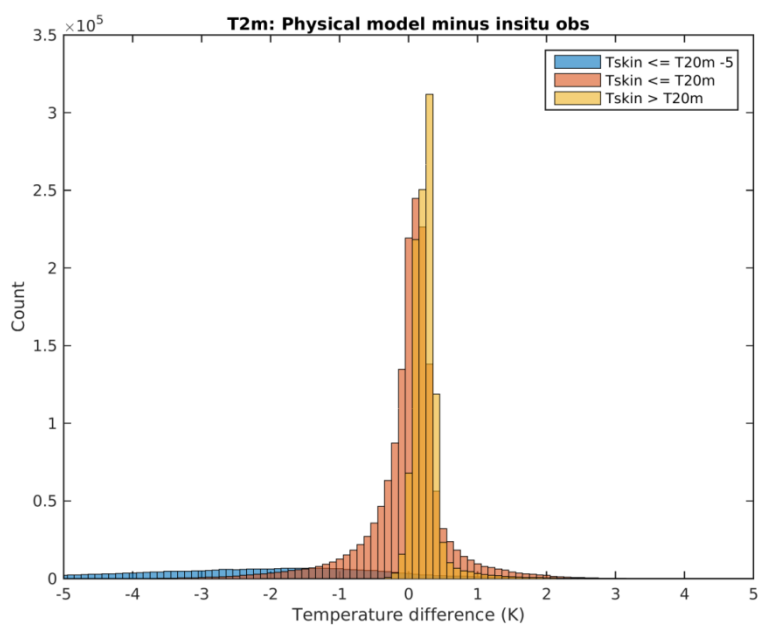
Figure 3.6.3: *Distributions of physically-derived minus observed T2m observations (°C) for unstable (yellow), stable (red) and very stable (blue) conditions for the Barrow Facility over snow covered grounds, 2003-2015.*

Table 3.6.2: *Summary of physical model results using in situ observations from the Barrow Facility over snow covered grounds, 2003-2015. Column 'Bias' Indicates the mean modelled minus observed T2m difference, 'STD' is the standard deviation, 'RMSD' is the root-mean square difference and 'r' is the correlation coefficient.*

	Bias	STD	RMSD	r	Count
Unstable conditions (Tskin > T1)	0.3°C	0.3°C	0.4°C	1.00	1 061 224
Stable conditions (T1 - 5°C < Tskin ≤ T1)	0.0°C	0.6°C	0.6°C	1.00	1 604 201
Very stable conditions (Tskin ≤ T1 - 5°C)	-2.3°C	2.1°C	3.1°C	0.96	320 686
All conditions	-0.2°C	1.1°C	1.1°C	0.99	2 986 111

Figure 3.6.4 shows time series of observed and modelled data from three selected one month periods, January, April and October 2007. The summer period was left out due to lack of snow, and the year 2007 was chosen because it has good data coverage. For these periods, sensitivity experiments were carried out on hourly data, similar to the experiments described in section 3.6.2, but for the observed Arctic conditions. Each of the (non-height)

input parameters was perturbed, keeping all other parameters fixed at each time step, and the coloured lines in Figure 3.6.4 show the ranges described by one standard deviation of each of the perturbed results. The temperatures *Tskin* and *T1* were perturbed linearly between ±5°C of the observed; *z0m* logarithmically between $1 \cdot 10^{-4}$ and 50 m; *W* linearly between ±5 m/s of the observed with a minimum of 0 m/s; and the boundary layer height linearly between 0 and 3000 m.

The results demonstrate that there is virtually no sensitivity to the boundary layer height and only a low sensitivity to *z0m* (in most cases less than 1°C), which is consistent with the results presented in section 3.6.2. For situations with very stable conditions, the results are sensitive to changes in wind speed, while for other conditions, the sensitivity to wind speed is low. During both stable and unstable conditions, the dependence on variations in *T1* is almost one-to-one, but for the very stable conditions, the dependence on variations in *T1* diminishes. The dependence on variations in *Tskin* is smaller, generally between ¼ and ½ of the dependence on *T1*, except for the very stable conditions, where the dependence becomes close to one-to-one.
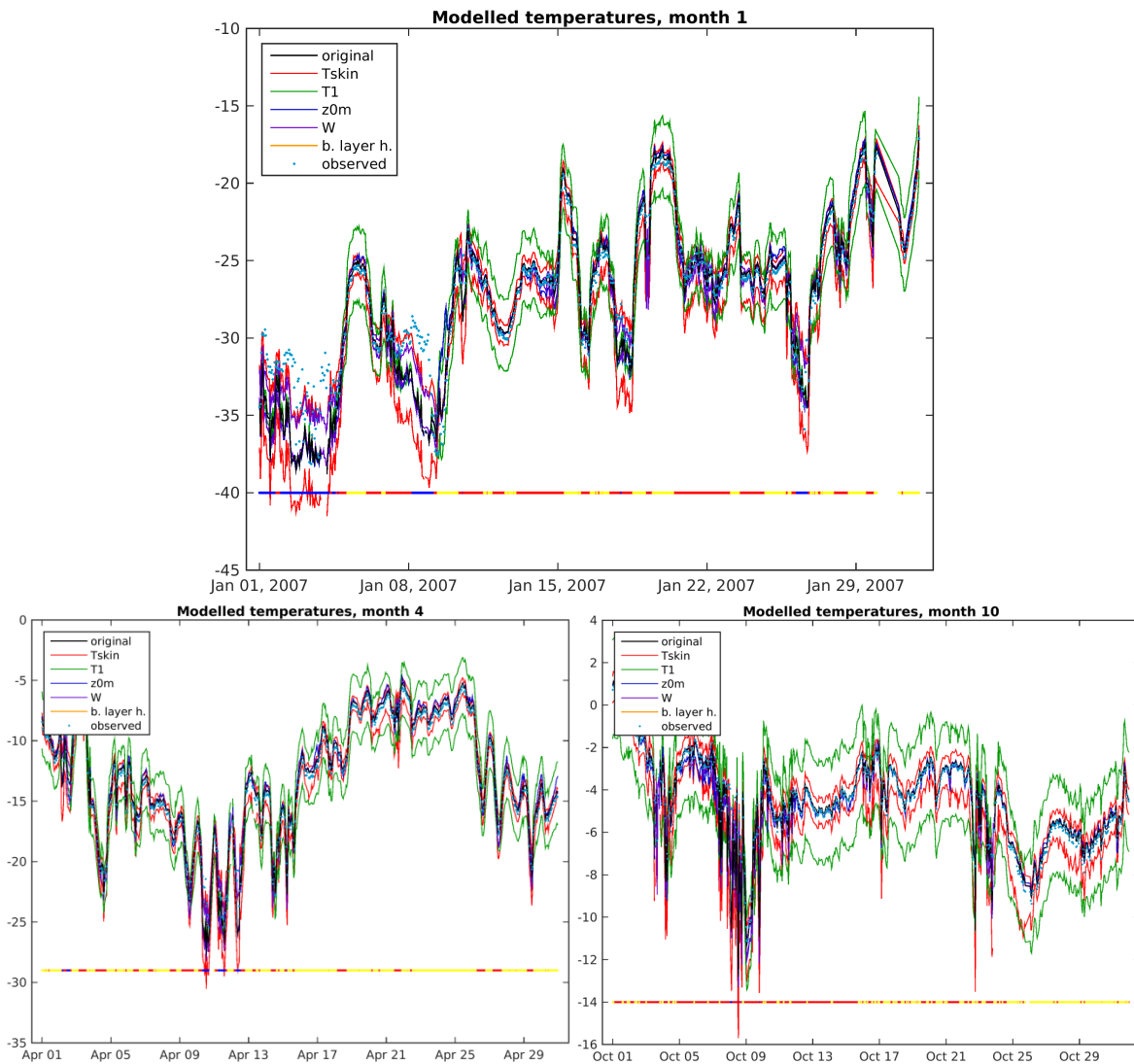
Figure 3.6.4: *Time series of modelled and observed T2m (°C) for testing model sensitivity from January (top), April (lower left) and October (lower right) 2007. The black line shows modelled temperature with the unperturbed model settings (using actual observations), cyan dots show observations, and the colored lines show ±1 STD for perturbation experiments with variation of Tskin (red), T1 (green), z0m (blue), W (purple), and boundary layer height (yellow). The bottom colour band indicate unstable (yellow), stable (red) and very stable (blue) conditions.*

This shows that when the physical model is applied to a combination of satellite observations of skin temperatures and upper air temperature and winds from atmospheric reanalysis models or weather forecasts, any errors in the upper air parameters will be propagated almost one-to-one into the modelled T2m. This is especially important in the Polar Regions, where atmospheric model results are of decreased quality compared to other regions of the world (see also Section 3.4).

3.6.5 Test application of the physical model over land using SEVIRI LST and ERA-Interim atmospheric data

The application of the physical model on a regional-to-global scale requires a consistent set of the input data required by the model. In this section, the feasibility of using the model in this way is assessed by calculating six-hourly *T2m* from satellite retrievals of *Tskin*, using reanalysis data to estimate lower atmospheric temperature (*T1*) and wind speed (*W*). The predicted *T2m* are evaluated by comparison to spatially and temporally collocated station observations of *T2m* from the HadISD data set (Dunn et al., 2012). The analysis is performed for January, April, July and October 2010.

The LST data are sourced from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) GlobTemperature 'GT_SEG_2P' products, which utilised the Land Surface Analysis Satellite Application Facility (LSA-SAF) LST retrieval scheme (Trigo et al., 2008; Freitas et al., 2010, Freitas et al., 2013) and for which uncertainty estimates were estimated in EUSTACE consistent with those estimated for other satellite data used elsewhere in the project. Atmospheric temperature and wind speed is obtained from the lowest model level (10m height) of the ERA-Interim reanalysis (Dee et al., 2011). The physical model is run for individual SEVIRI pixels, which have a resolution of 3 x 3 km at 0° latitude/longitude. Estimates for *z0m* for each pixel are obtained from the 300-m Land Cover Climate Change Initiative (CCI) data set (Bontemps et al., 2012, Poulter et al., 2015), based on the dominant land cover class within each SEVIRI pixel (Table 3.6.1). A fixed boundary layer height of 1000m is assumed for all T2m calculations as this information is unavailable in the ERA Interim NetCDF files available for the study. The impact of this assumption is expected to be negligible, given the apparent insensitivity of the modelled T2m to the boundary layer height demonstrated in sections 3.6.2 and 3.6.3.

Figure 3.6.5 shows histograms of the SEVIRI LST minus HadISD *T2m* observations for April 2010 over Europe, with separate panels for each six-hourly time slot. All distributions show pronounced 'cold tails', where LST is significantly colder than *T2m*. The cold tails indicate that the SEVIRI LST data suffer from significant cloud contamination, which causes LSTs to

be cold-biased, despite using the quality information present in the GlobTemperature LST files to exclude all pixels flagged as 'cloud contaminated' or 'low confidence' from the analysis. Fortunately, HadISD includes observations of cloud fraction (0 to 8 oktas), although the data availability is not as complete as for *T2m*. Figure 3.6.6 shows the same SEVIRI LST minus HadISD *T2m* observation distributions but only using satellite-station matchup pairs where the observed cloud fraction is reported to be 0 oktas. Although the number of data is significantly reduced, the distributions are now close to Gaussian, although there still appears to be some residual cloud contamination. This could be due to observer error in the HadISD data, or more likely because of cloud parallax effects, where cloud that is not directly above the HadISD station is obscuring the satellite view of the surface because of the satellite view geometry. As expected, the relationship between LST and *T2m* differs with time of day. LST is cooler at night (0Z, 18Z) and warmer during the day (6Z, 12Z), when clear-sky insolation usually elevates LST above *T2m* (Good, 2016 and Section 3.3).
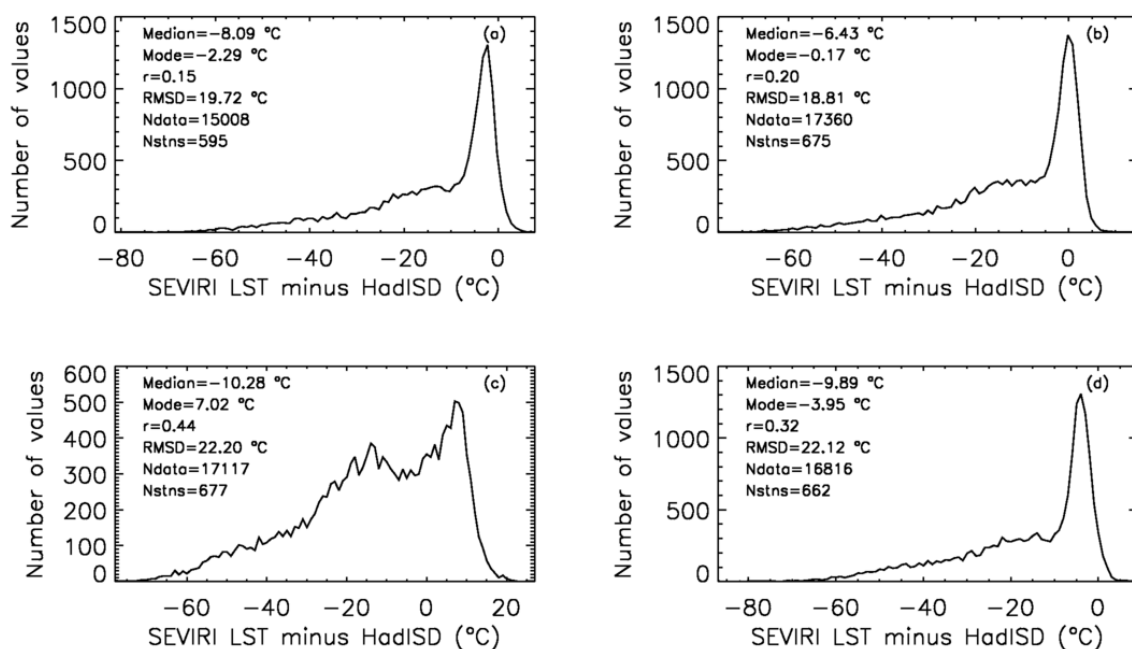

Figure 3.6.5: *Distributions of collocated SEVIRI LST minus HadISD T2m observations for (a) 0Z, (b) 6Z, (c) 12Z and (d)18Z over Europe during April 2010. 'r' is the correlation coefficient between SEVIRI and HadISD. The median and modal SEVIRI minus HadISD*

difference is indicated on the plot together with the Root Mean Square Difference (RMSD). 'Ndata' indicates the total number of matchups and 'Nstns' indicates the number of unique HadISD stations.
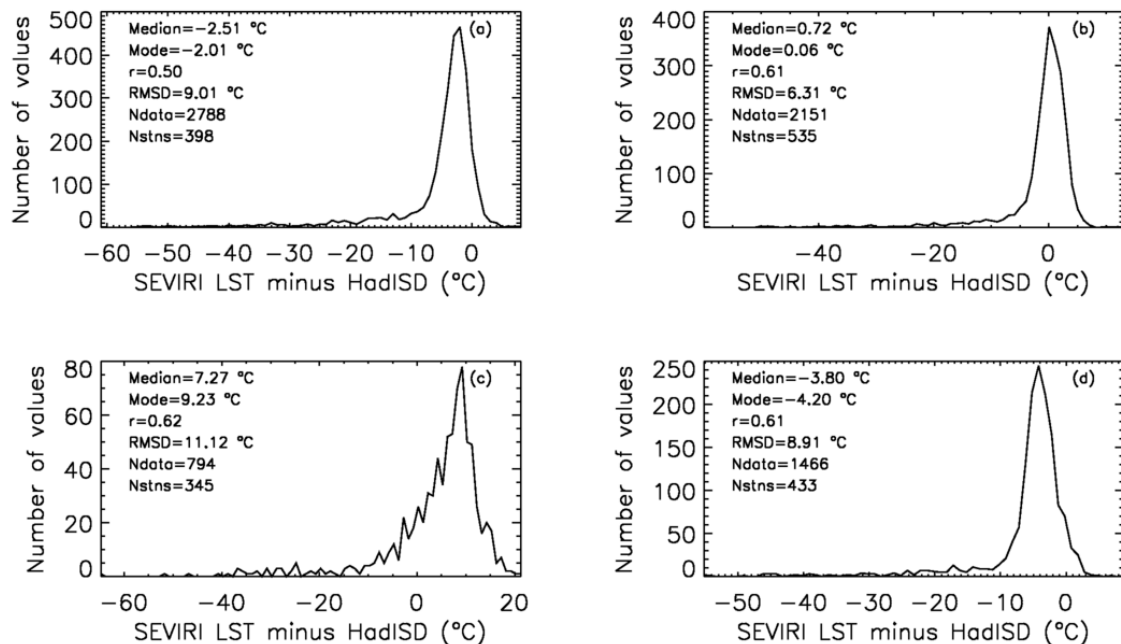

Figure 3.6.6: As for Figure 3.6.5 but for observations where the HadISD cloud fraction observation is present and recorded as zero oktas.
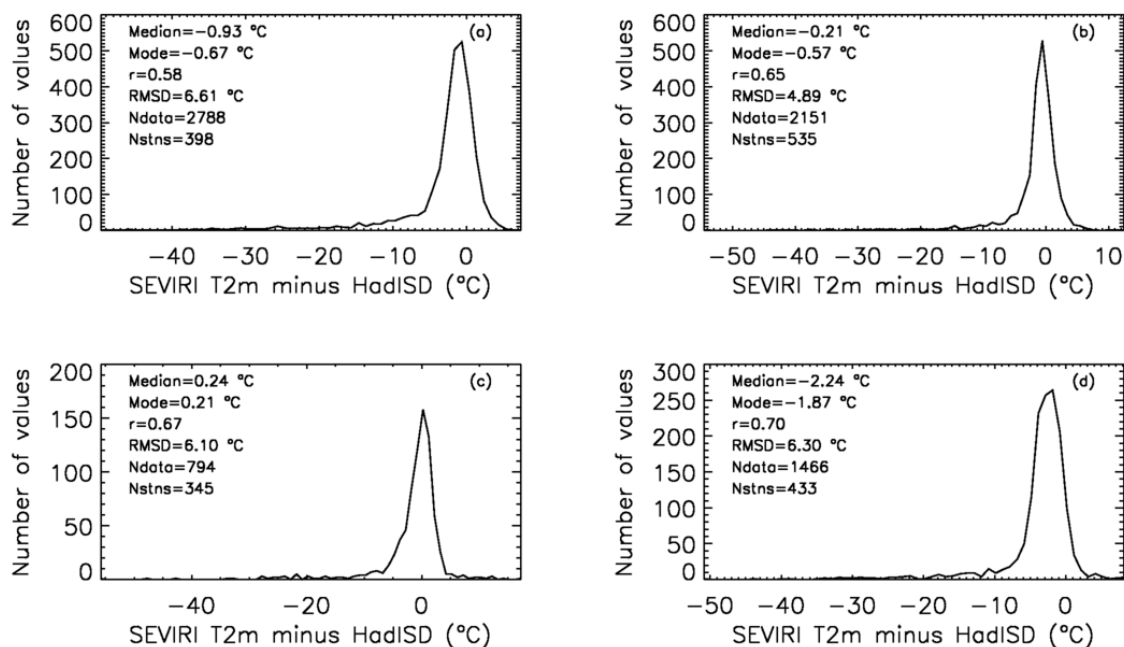
Figure 3.6.7: *As for Figure 3.6.5 but for predictions of T2m using the physical model with SEVIRI LST and ERA-Interim atmospheric parameters.*

Figure 3.6.7 shows the equivalent distributions and statistics for the comparison between the physically-derived *T2m* (using SEVIRI LST and ERA-Interim data) and HadISD *T2m* observations. In all cases, the median modelled minus observed difference is closer to zero and the root-mean square difference (RMSD) is reduced by a few °C, compared with the analysis using LST. It is worth noting that the RMSD will be inflated by the remaining cloud contamination. The correlation coefficients are also slightly improved compared with the LST-*T2m* comparisons in Figure 3.6.6.

Table 3.6.3 summarises the results for all months for Europe (35 to 70 °N, -10 to 20 °E), North Africa (0 to 35 °N, -15 to 15 °E), and South Africa (0 to -35 °N, 5 to 35 °E). These regions encompass most of the SEVIRI land field of view. Longitudinal ranges have been restricted to 30° in order to avoid very large variations in local solar time across a study region (30°= 2 hours variation in local solar time). Overall, the results are encouraging. Median modelled minus observed differences are typically a few tenths of a degree, and nearly all are <|1.0|. Median differences that exceed |1.0| are most common in the 18Z

results for Europe and North Africa, and are restricted to the 12Z results for South Africa. This suggests that the model may not be capturing the shape and/or timing of the *T2m* diurnal cycle accurately, due to the simplified model used. The RMSDs for Europe range between 4.9 and 9.0 °C, with a mean over all seasons/times of 6.1 °C. The RMSDs are lower for the African regions, with a mean of 4.0 and 3.4 °C for North and South Africa, respectively. These lower RMSDs are likely to be because residual cloud contamination will be higher over Europe owing to stronger cloud parallax effects at these latitudes (the SEVIRI view-zenith angle is lower close to the Equator so the satellite will 'see' cloud that is closer to that observed in situ). Correlation coefficients are moderate to high, tending to be higher in winter months and lower in the summer, with the mean values of 0.70, 0.81 and 0.71 for Europe, North and South Africa, respectively.

Table 3.6.3: *Summary of physical model results using SEVIRI LST. Column 'Med.' Indicates the median SEVIRI minus HadISD T2m difference. 'RMSD' is the root-mean square difference and 'r' is the correlation coefficient.*

Month	Time	Europe			North Africa			South Africa		
		Med. (°C)	RMSD (°C)	r	Med. (°C)	RMSD (°C)	r	Med. (°C)	RMSD (°C)	r
January	0Z	-0.8	6.3	0.84	-0.8	4.0	0.85	-0.1	6.4	0.34
	6Z	-0.5	5.2	0.89	-0.2	3.4	0.85	-0.4	5.5	0.36
	12Z	-0.6	5.7	0.87	0.0	1.8	0.96	1.9	3.3	0.83
	18Z	-1.3	5.0	0.89	-2.4	3.7	0.93	0.7	2.5	0.77
April	0Z	-0.9	6.6	0.58	-1.3	9.8	0.51	0.6	6.5	0.45
	6Z	-0.2	4.9	0.65	-0.4	4.2	0.79	0.2	1.8	0.89
	12Z	0.2	6.1	0.67	0.9	3.2	0.88	1.3	2.4	0.92
	18Z	-2.2	6.3	0.70	-1.7	5.2	0.83	0.3	2.9	0.77
July	0Z	0.0	7.1	0.55	-0.6	3.5	0.72	0.6	4.1	0.78
	6Z	-0.4	5.8	0.56	0.4	2.7	0.71	0.7	2.6	0.89
	12Z	-0.3	6.9	0.63	1.2	3.0	0.84	0.2	2.0	0.88
	18Z	-1.9	9.0	0.56	-1.2	2.5	0.92	-0.3	4.0	0.71
October	0Z	-0.3	6.5	0.61	-0.3	6.1	0.60	-0.1	4.3	0.68
	6Z	0.3	4.9	0.71	0.4	2.4	0.92	1.0	2.4	0.85
	12Z	-0.3	5.5	0.75	1.0	4.4	0.77	1.9	4.7	0.64
	18Z	-0.9	5.6	0.69	-1.1	3.2	0.91	-0.3	4.7	0.57
Average values		-0.6	6.1	0.70	-0.4	4.0	0.81	0.5	3.4	0.71

3.6.6 Test application of the physical model over oceans using CCI SST and ERA-Interim atmospheric data

The physical model described in the above section was used to estimate air temperatures over the ocean from sea surface skin temperatures based on satellite retrievals. The SST skin retrievals were taken from the SST CCI ATSR data set on a daily 0.25° latitude/longitude grid. Atmospheric boundary conditions – wind speed and air temperature at the lowest atmospheric level – were taken from the ERA-interim reanalysis (for 2005-2010). The roughness parameter was set to that of the sea (very smooth, 0.0002) and other parameters were set as for the land data. Four time steps are provided each day at 0000 0600 1200 and 1800. Four daily estimates of the 2m air temperature were made using the same SST skin temperature and these were averaged to form a daily mean.

The air temperatures estimated from satellite data using the physical model and the air temperatures estimated from satellite data using the empirical relationships were each compared to in situ data for the period 2005-2010. On average the estimates from the physical model had a cold bias when compared to the in situ measurements which is opposite to the expected bias. The 2m air temperatures output by the physical model should be warmer than the 10m air temperatures represented by the in situ data. In comparison, the empirical estimate had a slight warm bias. For the physical model, the magnitude of the cold bias exceeds 1°C across the Northern Hemisphere oceans, but in the tropics, the average bias was typically less than 0.5°C with some areas showing a small warm bias. For the empirical model, the tests showed that local biases did not exceed 0.5°C except in a few locations and were more consistent in sign.

The standard deviations of the differences between the estimated and measured air temperatures were very similar for the two different methods (physical model and empirical model): around 1.5°C. This implies that if the residual biases in the physical model can be dealt with (and their persistence suggests that this is viable) then the physical model can produce air temperatures that are as good as the empirical model derived here.

It is worth noting that the lowest atmospheric level in the reanalysis is notionally 10m which is the height to which the in situ ship data are harmonised. Comparing the in situ data at a height of 10m directly to the reanalysis at a height of 10m results in identification of very large apparent biases in the reanalysis. Given the sensitivity of our physically-derived air temperature estimates from satellite to our prescription of 10m air temperature (Section 3.6.2), this results in the cold bias we see in the estimated surface air temperature over the ocean.

3.6.7 Conclusions and outlook

This section presents the results of experiments to assess the feasibility of using a physical model to predict *T2m* from satellite *Tskin* data. In a project such as EUSTACE, this would provide an alternative source of satellite-derived *T2m* data to satellite-derived estimates based on surface-specific empirical statistical models. The advantage to such a scheme is a consistent approach over all surfaces (land, ocean, ice and lakes) and one that does not directly use station-based T2m observations to train a model, leaving these observations free to be utilised elsewhere, e.g. for validation. The main disadvantage of the physical model is the dependency on atmospheric data. For a EUSTACE-type analysis, the only viable source of these data is from reanalyses, which suffer from limitations and inaccuracies, particularly in regions where observations are sparse, such as at the poles. This leaves the surface air temperature estimates tied to the dynamical reanalysis, which removes the intended independence of EUSTACE estimates from NWP-model-based data. The physical model presented here is very sensitive to the upper air temperature so that any errors in this parameter will be propagated directly into the T2m estimate on a near one-to-one basis. Computational expense is also higher for the physical model compared with the empirical statistical models.

Nevertheless, this study has demonstrated that that the physical model tested here appears to produce T2m estimates that are comparable to T2m estimates based on empirical statistical models (also see sections 3.1-3.3). The accuracy of the physical model is good when used with high-quality in situ observations, except when the atmospheric stratification is very stable. This is true even for extreme Arctic conditions and a mean model-predicted minus observed T2m difference of -0.2 °C (RMSD 1.1 °C) is obtained for the test case presented here at the ARM-Barrow site in Alaska. Reasonable results are also obtained when trialling the model with inputs based on satellite LST and ERA-Interim reanalyses over Europe and Africa. Median model-predicted minus observed T2m differences are ~|0.5| °C, with an estimated RMSD of ~ 4 °C, although this can be much higher where the satellite LSTs are cloud-contaminated. Given the 1-3 °C accuracy of the satellite LST data used here (Trigo et al., 2008; Freitas et al., 2010, Freitas et al., 2013), inaccuracies in the ERA-Interim data, and the likely variance inflation due to the station point vs. satellite areal-averaged pixel matchup uncertainty, which may be ~ 2 °C over land (K. Veal, University of Leicester, Personal Communication) this is a very encouraging result and is comparable to the performance of the LSAT empirical statistical model (section 3.3).

The physical model predicts instantaneous *T2m* so using the model to estimate daily Tmin, Tmax or Tmean would be best achieved using high-frequency *Tskin* observations to capture

the T2m diurnal cycle. This could then be used to calculate Tmin and Tmax directly. For most of the earth, this could be achieved using geostationary *Tskin* observations. At high latitudes, where geostationary data are less reliable owing to the high satellite view angle, polar-orbiting data could be used, which provide more frequent overpasses at high latitudes. The *Tskin* data could be spatially aggregated to the required output *T2m* resolution before running the model to reduce computational expense. A reanalysis product, such as ERA-Interim would be needed to provide the atmospheric parameters required by the model. This would have to be temporally interpolated to the frequency of the *Tskin* observations; the impact of this has not been assessed in this study. The evidence presented in section 3.6.2 suggests that the physical model tested here may not capture the timing of the diurnal cycle accurately. However, for calculation of Tmin and Tmax from the diurnal cycle of predicted *T2m* this may not be an issue provided the amplitude is correct. As for the empirical statistical models, cloud contamination in satellite Tskin data remains the biggest challenge in estimating T2m from satellite data.

3.6.8 References

Atmospheric Radiation Measurement (ARM) Climate Research Facility. 2003, updated. Facility-specific multi-level Meteorological Instrumentation (TWR). 2003-10-31 to 2016-11-21, 71.323 N 156.609 W: North Slope Alaska (NSA) Central Facility, Barrow AK (C1). Compiled by J. Kyrouac and D. Cook. Atmospheric Radiation Measurement (ARM) Climate Research Facility Data Archive: Oak Ridge, Tennessee, USA. Data set accessed 2016-11-25 at http://dx.doi.org/10.5439/1025310.

Benali A, A. C. Carvalho, J. P. Nunes, N. Carvalhais, A. Santos (2012), Estimating air surface temperature in Portugal using MODIS LST data, Remote Sens. Environ. 124: 108–121, doi: 10.1016/j.rse.2012.04.024.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes, Geosci. Model Dev., 4, 677-699, doi:10.5194/gmd-4-677-2011, 2011.

Bontemps, S., M. Herold, L. Kooistra, A. van Groenestijn, A. Hartley, O. Arino, I. Moreau and P. Defourny (2012), Revisiting land cover observations to address the needs of the climate modelling community. Biogeosciences Discussions, 8(4), 7713-7740 (doi:10.5194/bg-9-2145-2012).

Chen, F., Y. Liu, Q. Liu and F. Qin (2014), A statistical method based on remote sensing for the estimation of air temperature in China. Int. J. Climatol, doi: 10.1002/joc.4113.

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., Huntingford, C., and Cox, P. M.: The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics, Geosci. Model Dev., 4, 701-722, doi:10.5194/gmd-4-701-2011, 2011.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F. (2011), The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q.J.R. Meteorol. Soc., 137: 553–597. doi:10.1002/qj.828.

Dunn, R. J. H., et al. (2012), HadISD: A Quality Controlled global synoptic report database for selected variables at long-term stations from 1973-2011, Clim. Past, 8, 1649-1679, 2012

Freitas, S. C., I. F. Trigo, J. M. Bioucas-Dias, and F. M. Göttsche (2010), Quantifying the uncertainty of land surface temperature retrievals from SEVIRI/Meteosat, IEEE Trans. Geosci. Remote Sens., 48(1), 523–534.

Freitas, S. C., I. F. Trigo, J. Macedo, C. Barroso, R. Silva, and R. Perdigão (2013), Land surface temperature from multiple geostationary satellites, Int. J. Remote Sens., 34(9–10), 3051–3068, doi:10.1080/01431161.2012.716925.

Good, E. (2015), Daily minimum and maximum surface air temperatures from geostationary satellite data, JGR-Atmospheres, doi:10.1002/2014JD022438.

Good, E. (2016), An in situ-based analysis of the relationship between land surface 'skin' and screen-level air temperatures, J. Geophys. Res. Atmos., doi:10.1002/2016JD025318.

Janatian, N., Sadeghi, M., Sanaeinejad, S. H., Bakhshian, E., Farid, A., Hasheminia, S. M. and Ghazanfari, S. (2016), A statistical framework for estimating air temperature using MODIS land surface temperature data. Int. J. Climatol.. doi:10.1002/joc.4766

Kilibarda, M., T. Hengl, G. B. M. Heuvelink, B. Gräler, E. Pebesma, M. Perčec Tadić, and B. Bajat (2014), Spatio-temporal interpolation of daily temperatures for global land areas at 1km resolution, J. Geophys. Res. Atmos., 119, 2294–2313, doi:10.1002/2013JD020803.

Lüpkes, C., Vihma, T., Birnbaum, G., & Wacker, U. (2008), Influence of leads in sea ice on the temperature of the atmospheric boundary layer during polar night, Geophysical Research Letters, 35(3).

Oyler, J. W., A. Ballantyne, K. Jencso, M. Sweet and S. W. Running (2015), Creating a topoclimatic daily air temperature dataset for the conterminous United States using homogenized station data and remotely sensed land skin temperature. Int. J. Climatol, 35, 2258–2279. doi:10.1002/joc.4127.

Parmentier, B., B. J., A. M. McGill, Wilson, J. Regetz, W. Jetz,, R. Guralnick, M.-N. Tuanmu, and M. Schildhauer (2015), Using multi-timescale methods and satellite-derived land surface temperature for the interpolation of daily maximum air temperature in Oregon, Int. J. Climatol., 35, 3862–3878, doi:10.1002/joc.4251.

Poulter, B., N. MacBean, A. J. Hartley, I. Khlystova, O. Arino, R. A. Betts, S. Bontemps, M. Boettcher, C. Brockmann, P. Defourny, S. Hagemann, M. Herold, G. Kirches, C. Lamarche, D. Lederer, C. Ottlé, M. Peters, and P. Peylin (2015), Plant functional type classification for Earth System Models: results from the European Space Agency's Land Cover Climate Change Initiative. Geosci. Model Dev. Discuss. 8(1), 429–462 (doi:10.5194/gmdd-8-429-2015)

Sun, Y. J., J. F. Wang, R. H. Zhang, R. R. Gillies, Y. Xue, Y. C. Bo (2005), Air temperature retrieval from remote sensing data based on thermodynamics, Theor. Appl. Climatol., 80(1), 37–48, doi:10.1007/s00704-004-0079-y.

Stull, R., 2015: "Practical Meteorology: An Algebra-based Survey of Atmospheric Science." Univ. of British Columbia. 938 pages. isbn 978-0-88865-176-1

Tetzlaff, A., Kaleschke, L., Lüpkes, C., Ament, F., & Vihma, T. (2013), The impact of heterogeneous surface temperatures on the 2-m air temperature over the Arctic Ocean under clear skies in spring, The Cryosphere, 7(1), 153-166.

Trigo, I. F., I. T. Monteiro, F. Olesen, and E. Kabsch (2008), An assessment of remotely sensed land surface temperature, J. Geophys. Res., 113, D17108, doi:10.1029/2008JD010035.

Vihma, T., & Pirazzini, R. (2005), On the factors controlling the snow surface and 2-m air temperatures over the Arctic sea ice in winter, Boundary-layer meteorology, 117(1), 73-90.

Vihma, T., Jaagus, J., Jakobson, E., & Palo, T. (2008), Meteorological conditions in the Arctic Ocean in spring and summer 2007 as recorded on the drifting ice station Tara, Geophysical Research Letters, 35(18).

Zhang, W., Z., Y. Huang , Y. Yu and W. Sun (2011), Empirical models for estimating daily maximum, minimum and mean air temperatures with MODIS land surface temperatures, International Journal of Remote Sensing, 32:24, 9415-9440, DOI: 10.1080/01431161.2011.560622.

4. Lessons learnt and future work

The use of empirical models for developing relationships between skin and near surface air temperature requires in situ observations for each surface type. The amount of in situ observations were particularly critical for the Polar Regions and for lakes. This was a known issue and efforts were allocated in WP3 to rescue and quality control observations from scientific campaigns. These rescuing efforts improved the situation for the Polar Regions, but the amount of available in situ observations from lakes turned out to be below the critical mass required, and another strategy had be pursued. It is evident that for future work, more accurate relationships could be obtained if more in situ reference data were rescued followed by an improved characterization of the different types of observations.

A common uncertainty approach was pursued within the different regions and satellite observations. This was a parallel track along with the development of relationships and had not been done before for the satellite observations over land and ice. The result from this work is a unique collection of satellite data sets with a consistent approach towards uncertainty estimation for all surfaces. It is, however, also clear, that modelling and verifying the different uncertainty components in the satellite retrievals are huge tasks that could be addressed in a future dedicated EU project. As a result, improvements can be made to these uncertainty components and in the understanding of the characteristic scales of the different components. In addition, it turned out to be challenging to communicate how to use these uncertainties to the users in WP2. Several separate meetings have already been held on this topic and more meetings are required when WP2 will start working with the data.

The results from the physical modelling show some promising results and have the potential to provide the next generation relationship method. Future developments outside this project could focus upon further utilization of the physical models. It is important to stress, however, that this approach depends upon the boundary conditions from NWP to a much higher degree than it depends upon the satellite observations. Using the physical modelling approach will thus reduce the new information gained from the satellite observations.

Finally, the consistency in the surface air temperature relationships at the boundaries between two surface types (edgelands) could be improved. This is particularly the case for the temporally and spatially varying marginal ice zone, but also applies for the other edgelands, such as the ocean – land interface.