Horizon 2020

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

EUSTACE

(GRANT AGREEMENT 640171)

EU SURFACE TEMPERATURE FOR ALL CORNERS OF EARTH

DELIVERABLE D3.2

Inter-comparison report for the intermediate fields inferred from retrieval

Deliverable Title	Inter-comparison report for the intermediate fields inferred from retrieval	
Brief Description	*This document reports the results of an inter-comparison of the EUSTACE Test Dataset and surface air temperatures from reanalyses.*	
WP number	3	
Lead Beneficiary	*Darren Ghent, University of Leicester*	
Contributors	*Karen Veal*	
Creation Date Version Number Version Date	12/11/2016 1 05/04/2018	
Deliverable Due Date Actual Delivery Date	M20 M40 (final)	
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**
0.1	12/11/2016	K L Veal	First draft
0.2	24/11/2016	K L Veal	Addressed comments
0.3	12/01/2017	K L Veal	Addressed further comments
0.4	14/03/2018	K L Veal	Addressed further comments
1	05/04/2018	N Rayner	Minor text changes, finalised for submission

EUSTACE

Table of Contents

Executive Summary

This document reports the findings of an inter-comparison of the EUSTACE Test Datasets (air temperature estimated from satellite data) with surface air temperatures from dynamical reanalyses. Differences in mean daily temperature are investigated by first regridding the datasets to a common grid and then calculating time series of daily mean hemispheric (northern and southern) temperatures, where both the reanalyses and the EUSTACE test datasets have data, and seasonal means of daily differences at the grid box level.

In some cases, good agreement was found between time series calculated from the EUSTACE Test Datasets and the ensemble median of the reanalyses. Some significant differences were apparent. All domains (land, ocean and ice) showed seasonally varying differences in the hemispheric means although this was less evident for northern hemisphere sea ice. The EUSTACE Land Test Dataset was generally colder than the ensemble median in the northern hemisphere and also showed negative differences of several kelvin over mountainous regions, which is not entirely expected from our validation against independent in situ data. The EUSTACE Ocean Test Dataset was generally warmer than the ensemble median by 0.5 to 1.0 K. These differences are larger than the differences found between EUSTACE test data and nighttime in situ data, in previous work and requires further investigation. At the western boundary currents the EUSTACE dataset was colder than the ensemble median by 1 to 3 K depending on location and season. Over land ice and sea ice, the difference in hemispheric mean was within the combined ±2σ uncertainty range for nearly all the time series (2007-2008). The largest differences over ice were in high altitude regions such as the Greenland ice sheet and parts of Antarctica and are likely due to warm biases in the reanalyses.

As these results relate to early test versions of the EUSTACE data, they may not be applicable to the final products. Later Deliverable reports will assess the performance of the final EUSTACE products.

1. Datasets

1.1. EUSTACE Test Dataset

This analysis focuses on the second version of the EUSTACE Test Dataset. The dataset includes estimates of air temperature from satellite retrievals over land, ocean and ice (note the first test dataset did not include estimates over ice). Data are provided on equal angle longitude-latitude grids with a spatial resolution of 0.25°.

1.1.1 EUSTACE Land Test Data

The EUSTACE Land Test Data provide estimates of daily minimum (Tmin) and daily maximum (Tmax) surface air temperature over land in daily files for the period starting 04/07/2002 and ending 12/04/2012. Air temperature was estimated from the MODIS Aqua skin temperature retrievals via a regression-based method. Data are provided on equal angle longitude-latitude grids with a spatial resolution of 0.25°. Uncertainties on the temperature estimates are provided at the grid box level. Three components are given: random, correlated and systematic. No length or time scales are given for the correlated component in this test version.

1.1.2 EUSTACE Ocean Test Data

EUSTACE Ocean Test Data estimates of mean daily marine air temperature (Tmean) are delivered in daily Level 3 files at a spatial resolution of 0.25° for 21/05/2002-08/04/2012. The estimates were derived from retrievals of SST at 0.2 m depth from the Along-track Scanning Radiometers. Uncertainty estimates are provided as ten separate components: a random component, two locally correlated components, two systematic components and five components associated with the Fourier coefficients of the model. No correlation scales are given in the files for this test version. However, advice was obtained from WP2 on the likely correlation scales of the different components. The correlation length and time scales for the first and second locally correlated components respectively are 100 km, 1 day and 1000 km, 5 days. The components associated with the Fourier coefficients are systematic in time and uncorrelated in space.

1.1.3 EUSTACE Ice Test Data

The EUSTACE Ice Data are daily estimates of Tmin, Tmax, and Tmean. The data provided cover 2007 and 2008. The separate components of the uncertainty estimates are made

available; random, locally correlated, large-scale systematic, and cloud mask uncertainty components are given with associated correlation length and time scales. The dataset also provides masks for land ice and sea ice (sea ice concentration above 30%). Masks are also provided which separate the sea ice into high concentration sea ice (above 85%) and the marginal ice zone (sea ice concentration between 30% and 85%). This test version of the daily estimates are based on a 24 hour period from 00:00 to 24:00 Universal Time unlike the land and ocean datasets which use a local solar day.

1.2. Reanalyses

For this study, five readily available state of the art dynamical reanalysis datasets have been chosen for inter-comparison with the EUSTACE Test Dataset. These are: ERA Interim [Dee et al., 2011] and ERA 20C [Poli et al., 2013], both produced by ECMWF; the NCEP/NCAR Reanalysis 1 [Kalnay et al., 1996]; NASA's Modern-Era Retrospective Analysis for Research and Applications, MERRA, [Rienecker et al., 2011]; and the Japanese 55 year Reanalysis, JRA55, from the Japanese Meteorological Agency [Kobayashi et al., 2015]. The sources for these datasets are provided in Section 4. All the reanalyses have global spatial coverage. The temporal coverage, temporal and spatial resolutions of the different reanalyses are given in Table 1-1.

Table 1-1: Temporal coverage and temporal and spatial resolutions of the reanalysis data used in the inter-comparison

Dataset	Temporal Coverage	Temporal resolution	Spatial Resolution (longitude x latitude)
ERA Interim	1979 - present	6 hourly	0.75° x 0.75°
ERA 20C	1900 - 2010	3 hourly	1.0° x 1.0°
JRA55	1958 - present	6 hourly	1.5° x 1.5°
MERRA	1979 - present	hourly	0.5° x 40'
NCEP/NCAR Reanalysis I	1948 - present	6 hourly	2.5° x 2.5°

2. Approach

Estimates of mean daily temperature (Tmean) from the EUSTACE Test Dataset and a median of an ensemble of reanalyses are compared on a common grid. Time series of mean hemispheric (north and south) Tmean over each of the ice, land, and ocean domains are presented in Section 3 along with maps of seasonal EUSTACE minus ensemble median difference. The rest of this section provides further details of the approach.

2.1. Calculation of the daily mean temperature

Calculation of an accurate minimum or maximum daily temperature from the reanalyses is unfeasible because of their coarse temporal resolutions. We, therefore, compare the EUSTACE daily average temperature with an average temperature from reanalyses that is calculated by simply, at a given grid point on the native grid, averaging the temperatures at different time points within the day. Separate grids were calculated using local solar time (for comparison with EUSTACE land and ocean estimates) and using universal time (for comparison with EUSTACE ice estimates).

The EUSTACE Ice and Ocean Test Datasets contain estimates of Tmean. However, the EUSTACE Land Test Dataset provides Tmin and Tmax, not Tmean, so Tmean was calculated over land by simple averaging of the estimates of Tmin and Tmax. In the intercomparison reported here, the uncertainties on Tmin and Tmax were assumed independent and so added in quadrature to obtain the uncertainty on Tmean. However, these uncertainties are not independent as the nighttime and daytime LSTs are used to estimate both Tmin and Tmax. The uncertainty on the Tmean is, therefore, underestimated here. This issue will be addressed for the intercomparison of the final EUSTACE dataset with reanalyses.

2.2. Regridding

The NCEP/NCAR Reanalysis 1 data have the lowest spatial resolution of 2.5° in longitude and latitude. This was chosen as the common grid for the analysis.

The reanalysis data were regridded onto the 2.5° grid by first interpolating the corner grid box values to centre grid box values, and then calculating a simple average of the native resolution grid boxes corresponding to the 2.5° grid box.

The EUSTACE estimates of Tmean were regridded to daily, 2.5° grids by simple averaging. The EUSTACE Test Dataset provides uncertainty estimates for each grid box separated into components according to the nature of their correlation (in space and time) with the uncertainty estimated for other grid boxes. This enables propagation of the uncertainty estimates through an averaging process.

In each dataset, uncertainty components are considered to arise from either random (not correlated), locally correlated on some length and time scale, or systematic (correlated on all length and time scales) effects. The EUSTACE ice dataset provides, for the locally correlated component, estimated length and time scales of 500 km and 5 days respectively. These are approximate synoptic scales and in this analysis are also adopted for propagation of the land locally correlated uncertainty estimates whilst recognizing that other scales may be more appropriate for this domain. In the final EUSTACE dataset, over land, the locally correlated component will be separated into surface and atmospheric components, reflecting the differing correlation scales of the individual components of uncertainty on the input LST.

In the case of the ocean uncertainty estimates, the two locally correlated components have different length and time scales: 1 day and 100 km for component C1 and 5 days and 1000 km for component C2.

The length of a 2.5° grid box is approximately 278 km at the equator, within the correlation length scales given for all the locally correlated components of uncertainty except ocean component C1. Here we treat all locally correlated components as correlated on propagation to the 2.5° grid. This method slightly overestimates the ocean component C1 because it is correlated on a smaller scale of 100 km. In the intercomparison of the final EUSTACE dataset and reanalyses, this component will be propagated first to 100 km (~ 1 degree) spatial resolution accounting for correlations and then to the 2.5° resolution. The random components and the ocean Fourier coefficient components are treated as uncorrelated; the systematic components are, of course, treated as correlated. The uncertainty on the 2.5° grid box average temperature, $\sigma_{2.5^\circ}$ for land and ice is therefore given by:

$$\sigma_{2.5^\circ}^2 = \frac{1}{N^2}\left(\sum_{i=1}^{N}\sigma_{i,R}^2 + \sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{i,C}\sigma_{j,C} + \sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{i,L}\sigma_{j,L} + \sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{i,Clc}\sigma_{j,Clc}\right),$$

where $\sigma_{2.5^\circ}^2$ is the uncertainty on the 2.5° gridbox mean, N is the number of 0.25° pixels that contribute to the mean, σ_R is the random uncertainty component, σ_C is the synoptic scale correlated component, and σ_L is the large-scale correlated component on the ith 0.25° pixel.

A separate component of uncertainty due to the cloud mask, σ_{Cld}, is only currently provided with the ice data; information in the test datafile advises that this component is systematic and that is how it is treated in this work.

For ocean, the 2.5° uncertainty is:

$$\sigma_{2.5°}^2 = \frac{1}{N^2}\left(\sum_{i=1}^{N}\sigma_{i,R}^2 + \sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{i,C1}\,\sigma_{j,C1} + \sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{i,C2}\,\sigma_{j,C2} + \sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{i,L1}\,\sigma_{j,L1} + \sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{i,L2}\,\sigma_{j,L2} + \sum_{i=1}^{N}\sigma_{i,P1}^2 + \sum_{i=1}^{N}\sigma_{i,P2}^2 + \sum_{i=1}^{N}\sigma_{i,P3}^2 + \sum_{i=1}^{N}\sigma_{i,P4}^2 + \sum_{i=1}^{N}\sigma_{i,P5}^2\right)$$

Here, σ_{C1} and σ_{C1} are the two locally correlated components of the uncertainty estimate, σ_{L1} and σ_{L2} are the two systematic components, and σ_{P1}, σ_{P2}, σ_{P3}, σ_{P4} and σ_{P5} are the five components of the uncertainty for the Fourier parameters.

2.3. Analysis

The analysis compares EUSTACE estimates of air temperature with air temperature given by the median of an ensemble of four or five reanalyses. The EUSTACE Ice estimates are compared with an ensemble of the five reanalyses described in Section 1.2. The ERA 20C does not cover the full period of the EUSTACE land and ocean datasets and is excluded from the ensemble for inter-comparison with these datasets.

After regridding of the datasets, time series of northern and southern hemispheric 5 day mean temperature were calculated from the EUSTACE Test Datasets and the reanalyses. In addition, seasonal means were calculated at the 2.5° grid box level. Grid boxes where the EUSTACE data contained fewer than twenty out of the possible one hundred 0.25° pixels were excluded from the analysis in all cases. This was done to reduce the effects of low sampling due to cloud or domain boundaries. In the case of both the hemispheric and seasonal means only coincident reanalysis and EUSTACE 2.5° grid boxes contributed to calculation of the means. In all cases, at least twenty valid days were required for calculation of a seasonal average.

Throughout this analysis the ensemble robust standard deviation (r.s.d.) is calculated from the median absolute deviation (m.a.d.) by multiplying by a scale factor which assumes that the range ±m.a.d. covers 50% of the normal cumulative distribution function that would best fit the data if outliers are excluded [Huber and Ronchetti, 2009]:

$$\sigma_{Ens} = 1.4826 \times m.a.d.$$

The uncertainty on the EUSTACE estimate is propagated through the calculation of the hemispheric and seasonal means. In the case of the seasonal mean temperature at a 2.5° grid box, the locally correlated component, σ_C, has a correlation timescale of five days. Those values within a five day window are considered correlated. The uncertainty on the seasonal mean at a 2.5° grid box, for land and ice, is, therefore,

$$\sigma^2_{Seasonal,2.5°} = \frac{1}{N^2}\left(\sum_{i=1}^{N}\sigma^2_{i,R} + \sum_{i=1}^{N}\sigma^2_{i,C} + 2\sum_{i=1}^{N-1}\sigma_{i,C}\sigma_{i+1,C} + 2\sum_{i=1}^{N-2}\sigma_{i,C}\sigma_{i+2,C} + \sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{i,Cld}\sigma_{j,Cld}\right)$$

Here, the indices i and j, denote days in the time series of data for a given grid box and season. The contribution of the locally correlated component is contained within the second, third and fourth terms inside the brackets. The second and third terms contain the correlations between uncertainties from days that are one and two days apart, respectively (we have taken advantage of the symmetry of the correlation matrix and doubled each term as the correlation of σ_i with σ_{i+1} is equal to the correlation of σ_{i+1} with σ_i). The fourth term in the brackets, the cloud mask term, is only included for the EUSTACE ice data.

In the case of the EUSTACE ocean test data, the first locally correlated component is considered uncorrelated (its correlation time scale is 1 day) and the second locally correlated component is again correlated on a five day time scale. The Fourier coefficient components are systematic in time. The uncertainty on the seasonal mean is, therefore

$$\begin{aligned}\sigma^2_{Seasonal,2.5°} = \frac{1}{N^2}\Bigg(&\sum_{i=1}^{N}\sigma^2_{i,R} + \sum_{i=1}^{N}\sigma^2_{i.C1} + \sum_{i=1}^{N}\sigma^2_{i,C2} + 2\sum_{i=1}^{N-1}\sigma_{i,C2}\sigma_{i+1,C2} + 2\sum_{i=1}^{N-2}\sigma_{i,C2}\sigma_{i+2,C2}\\ &+ \sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{i,L1}\sigma_{j,L1} + \sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{i,L2}\sigma_{j,L2} + \sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{i,P1}\sigma_{j,P1} + \sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{i,P2}\sigma_{j,P2}\\ &+ \sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{i,P3}\sigma_{j,P3} + \sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{i,P4}\sigma_{j,P4} + \sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{i,P5}\sigma_{j,P5}\Bigg)\end{aligned}$$

and again the indices i and j, refer to days in the time series for a given grid box and season.

In the case of the hemispheric means, the number of grid boxes is large and the positional relationships of neighbourhood correlations is two dimensional (only one dimension need be considered when dealing with correlations in time). A simpler approach is taken where the 2.5° grid boxes are, first, averaged to a grid resolution approximately equivalent to the correlation length scale of the locally correlated component and the hemispheric means calculated by averaging the lower resolution grid boxes in a second step. This method necessarily neglects local correlations across the edges of the lower resolution grid boxes. In the case of land data, a 5° x 5° grid is used which is approximately equivalent to 500 km at the equator and for ocean data a 10° x 10° grid, approximately 1000 km, is used. A 10° x 10° grid is used for the ice as all data are polewards of 60° latitude and 10° of longitude is approximately equivalent to 500 km. The locally correlated component is treated as correlated during this step. For example, for ice data, the uncertainty on a daily, 10° grid box temperature is

$$\sigma_{10^\circ}^2 = \frac{1}{N^2}\left(\sum_{i=1}^{N}\sigma_{i,R,2.5^\circ}^2 + \sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{i,C2.5^\circ}\sigma_{j,C,2.5^\circ} + \sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{i,L,2.5^\circ}\sigma_{j,L,2.5^\circ} + \sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{i,Cld,2.5^\circ}\sigma_{j,Cld,2.5^\circ}\right)$$

where the sum is over the individual 0.25° cells making up the 10° box.

The hemispheric daily means are then calculated from the data at the coarser resolution, the locally correlated component being treated as uncorrelated for this step. During the latter step the temperatures and the uncertainties are weighted according to the number of 2.5° valid grid boxes (w_j) available in each lower resolution box (j) and so, for example for ice, the uncertainty on the daily, hemispheric mean is

$$\sigma_{Hemisphere}^2 = \frac{1}{\left(\sum_{j=0}^{N} w_j\right)^2}\left(\sum_{i=1}^{N} w_i^2\sigma_{i,R,10^\circ}^2 + \sum_{i=1}^{N} w_i^2\sigma_{i,C,10^\circ}^2 + \sum_{i=1}^{N}\sum_{j=1}^{N} w_i\sigma_{i,L,5^\circ}w_j\sigma_{j,L,10^\circ} + \sum_{i=1}^{N}\sum_{j=1}^{N} w_i\sigma_{i,Cld,10^\circ}w_j\sigma_{j,Cld,10^\circ}\right)$$

Five day hemispheric mean temperatures are calculated at each daily time point by averaging the daily means for each 5 day period. The uncertainty components are

propagated according to their correlation time scales. For ice and land the locally correlated component and the systematic component are assumed correlated in the five day averaging. For ocean, the two systematic components, the five Fourier coefficient components and the second locally correlated component are all treated as correlated on a five day time scale. The first locally correlated component is treated as uncorrelated.

Time series of five day mean hemispheric temperatures are also calculated from each reanalysis. The ensemble median and r.s.d. is then calculated at every time point.

The uncertainty on the EUSTACE minus ensemble median difference is estimated by adding, in quadrature, the EUSTACE uncertainty and the ensemble r.s.d.

Note: In this report a correlation time-scale (τ) of 5 days has been interpreted as meaning a daily uncertainty on day d is correlated with other daily uncertainties for days within a 5 day window, $d-2$, $d-1$, d, $d+1$, and $d+2$, and not correlated with uncertainties outside the window. Since the correlation coefficient is estimated to decrease by a factor of $1/e$ for each 5 day increase in time interval, a better estimation of the uncertainty would result if the uncertainty on day d were assumed correlated within a 9 day window *i.e.* from $d-4$ to $d+4$. The uncertainties on the EUSTACE monthly and seasonal averages have therefore been slightly underestimated in this report. In future work, the window for correlation of uncertainty components in time will be estimated as $2\tau-1$.

3.Results

This section compares time series of northern and southern hemispheric five day mean temperature and seasonal mean temperature at the 2.5° grid box level from EUSTACE data and a median of an ensemble of reanalyses over land, ice, and ocean domains. The standard meteorological seasons are used, i.e. in the northern hemisphere, winter is December, January and February (DJF), spring is March, April and May (MAM), summer is June, July and August (JJA) and autumn is September, October and November (SON).

3.1. Land

In the northern hemisphere, EUSTACE test dataset spatial coverage varies seasonally with the lowest number of populated grid boxes in winter and the highest in summer (Figure 3-1). There is very little data above 60°N in the winter months (DJF). Low sample numbers are probably due to persistent cloud and will be exacerbated by our insisting on 20% of 0.25° cells being present in a grid box before inclusion of the grid box in the seasonal average. The EUSTACE hemispheric mean is consistently colder than the reanalysis ensemble median with a seasonally-varying discrepancy which is approximately -0.5 K in winter and around -2 K in summer. In the summer months, the discrepancy exceeds the 2 sigma combined (EUSTACE and reanalysis ensemble) uncertainty.

In the southern hemisphere, sampling numbers are lowest in the austral summer, again due to cloud affecting the satellite data (the cloud of the Inter-Tropical Convergence Zone affects a large proportion of the southern hemisphere land mass in this season). Also, there is no data over the Antarctic which may be due to overzealous post-processing cloud clearing; this extra cloud clearing step is used to reduce cloud contamination of the input LSTs. The dynamical reanalyses are more widely spread during the months of NDJ in the southern hemisphere and consequently the reanalysis ensemble r.s.d. is larger. There is again a seasonal variation in the hemisphere mean EUSTACE minus reanalysis ensemble difference with positive differences during the austral summer (DJF) reaching 0.5 K in January in most years and negative differences of -1.0 to -1.5 K in the austral winter months. However, the difference is within the 2 sigma uncertainty limits for nearly all points in the time series.

The regions that show the largest differences between the EUSTACE estimates and the reanalysis ensemble (Figure 3-2) are the mountainous regions of the Rockies and Sierra

Madre and the Andes in the Americas and the Himalayas in Asia. The Sahara and Arabian Peninsula show large differences in MAM and JJA. This may be due to the effects of dust on the satellite retrievals. The Brazil Highlands and a region of Africa including Angola and parts of the Congo also show large differences in JJA and SON; these are semi-arid regions and the seasonality of the differences matches their wet/dry seasons. Large areas of northern Russia show large differences in summer and also significant (> 2σ) differences in spring and autumn. Interestingly, the areas where we would expect problems due to cloud contamination of the satellite data, such as the tropical forests, show smaller differences.

It is instructive to compare these differences to the results of the validation of the EUSTACE test data against reference in situ data reported by Veal et al. (2016). In general, over the US, differences from reference in situ data are larger over the mountainous regions of the western US and smaller over the low-lying eastern US, particularly for Tmax. Over the western US, differences from the reference data (EUSTACE test data minus in situ) tend to be positive for Tmax for all seasons, except at the highest altitudes where they can be strongly negative. The same is true of Tmin, but the differences are smaller. Discrepancies between Tmean and the reanalysis ensemble median (EUSTACE minus reanalysis) are negative over the whole region. This suggests that the reanalysis ensemble median is biased warmer than the EUSTACE test data over this region at lower altitudes. In the coastal regions of the north eastern US, around New York state and New England, differences from the in situ reference data also tend to be positive, over the whole year in the case of Tmin and in SON and DJF in the case of Tmax. Here too, discrepancies between Tmean in the EUSTACE test dataset and the ensemble median are negative in SON and DJF, indicating that the reanalysis median may be more strongly positively biased in these seasons in this region. Over the greater part of the eastern US, the validation results are more consistent with the differences seen here relative to the reanalysis ensemble. JJA is the only season here where the discrepancy between the EUSTACE test dataset and the reanalysis is noticeably larger than between the EUSTACE test data and the in situ reference data. Later, the EUSTACE products will be validated against independent reference data over other parts of the world, so we will be able to determine more generally when and where the accuracy of the EUSTACE products is better or worse than that of the dynamic reanalysis median.

Note, the inter-comparison analysis described here compares a EUSTACE daily mean temperature calculated by averaging the minimum and maximum daily temperatures with an ensemble of mean temperatures from reanalyses calculated by averaging the temperatures

at all the available time points in the day. Some differences would therefore be expected. The MERRA analysis which gives hourly temperatures will be used to quantify the expected differences between the two methods of calculating Tmean in a later EUSTACE activity.

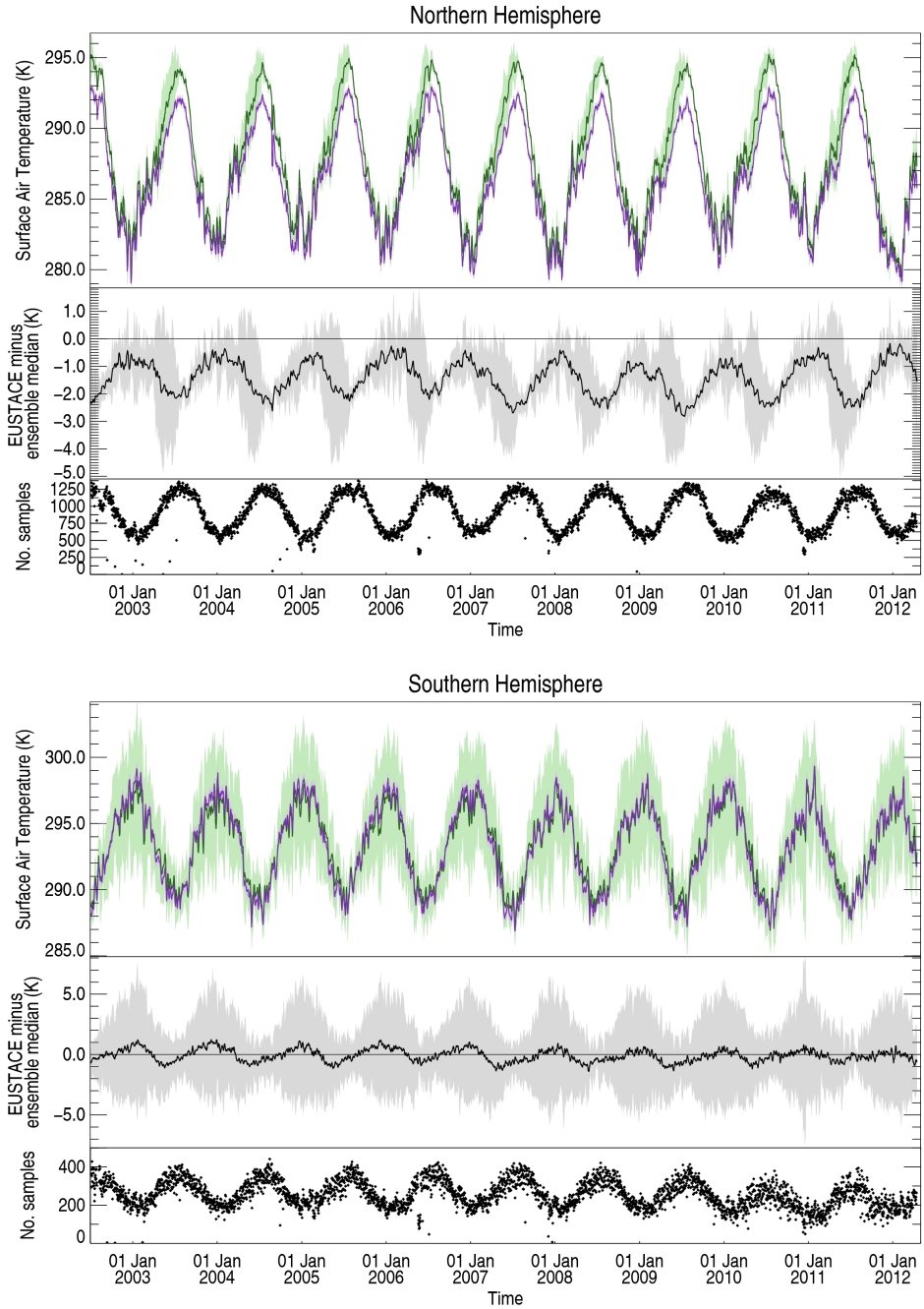


Figure 3-1:Time series of daily northern (top) and southern (bottom) hemisphere mean temperature (K) over land, 2002-2012. In each panel, the reanalysis ensemble median (green line) and test EUSTACE mean regional air temperature estimates are plotted at the top. The middle panel shows the EUSTACE minus reanalysis ensemble median daily difference and the bottom plot shows the number of coincident 2.5° gridboxes that contribute to the hemispheric mean. In the top panel, the green shaded area is the reanalysis ensemble median +/- twice the r.s.d. and the purple shaded area is the EUSTACE mean +/- the EUSTACE 2σ uncertainty. In the

middle panel, the grey shaded region is +/- twice the combined reanalysis ensemble r.s.d and EUSTACE uncertainty.

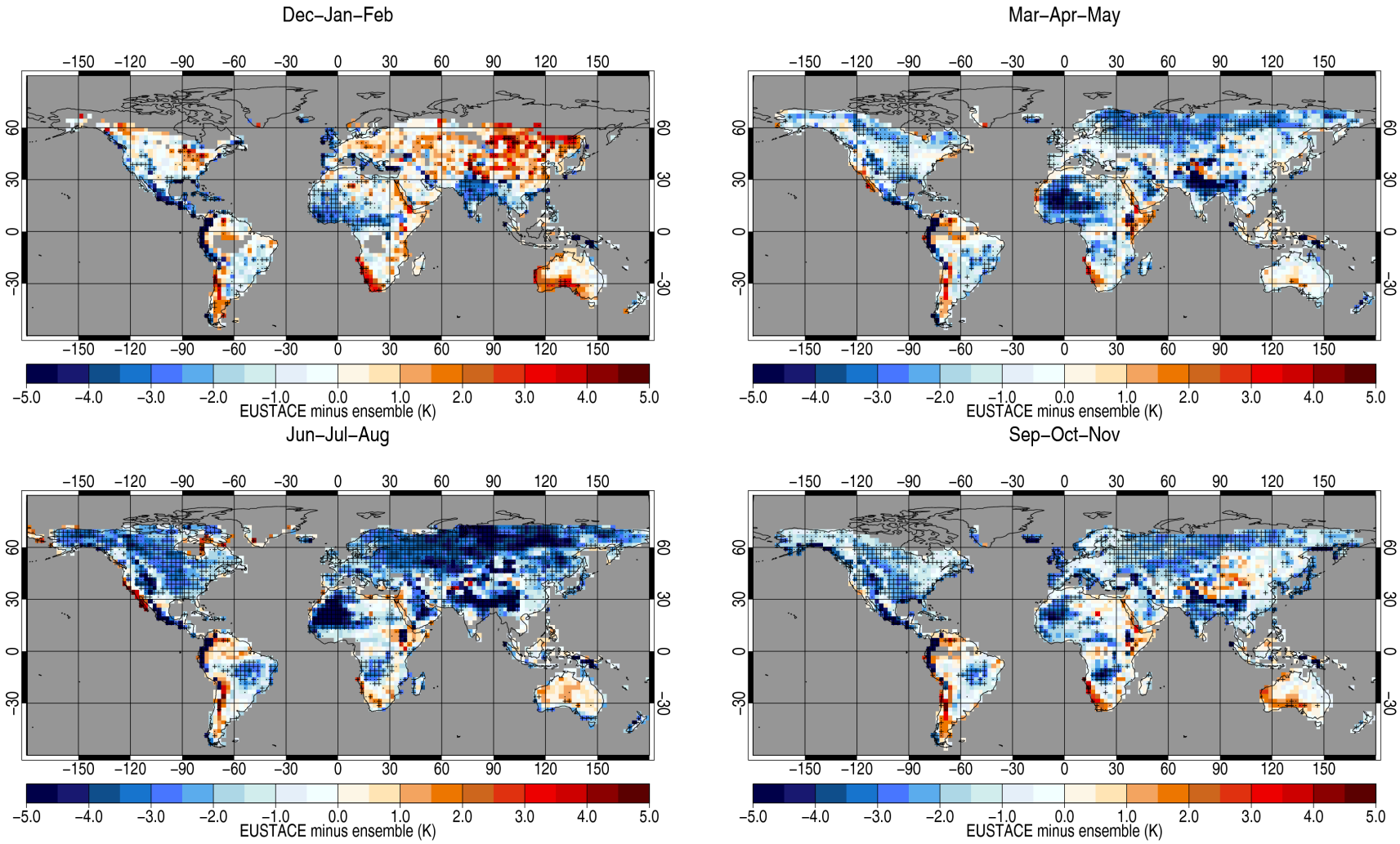


Figure 3-2: Seasonal mean EUSTACE minus reanalysis ensemble median daily difference in surface air temperature (K), averaged over the period 2002-2012. Differences that exceed the 2 (3) sigma uncertainty are marked with a dot (cross).

3.2. Ocean

The time series of northern hemisphere mean daily marine air temperature from the EUSTACE test dataset shows good agreement with the reanalysis ensemble median (Figure 3-3). There is a seasonally varying difference which varies between near zero in the boreal late spring/summer and reaches around 0.6 K in the late autumn/winter with EUSTACE warmer than the reanalysis ensemble median. This difference only exceeds the 2 sigma uncertainty for a few days each year, usually during summer.

In the southern hemisphere, the EUSTACE test dataset is systematically warmer than the reanalysis ensemble median with the difference varying from around 0.5 K in the summer (DJF) to about 0.8 K in the winter.

The largest differences appear at the extremes of the spatial coverage of the EUSTACE test data (high and low latitudes) where sample numbers are low and in coastal regions where the reanalysis grid box temperature includes air temperatures over land but the EUSTACE ocean test dataset does not (see Figure 3-4). Also, the relatively small northern hemisphere mean differences hide areas of compensating relative differences with large differences apparent at the western boundary currents where the EUSTACE test dataset is colder than the reanalysis ensemble median; the EUSTACE data are warmer elsewhere. Differences between the EUSTACE test estimates and the reanalysis ensemble are expected in the regions of the boundary currents. Temperature gradients are high in the regions of the boundary currents and the high gradients are less well resolved by the analyses which have lower spatial resolutions than the EUSTACE data. Also, these regions have a high frequency of cloud and cloud contamination of the input SST data would cause the EUSTACE test data to be biased cold. The large negative differences seen in the North Pacific (DJF and MAM) may also be caused by contamination of the SST with cloud or possibly of sea-ice. By contrast, the southern hemisphere is generally warmer in the EUSTACE test dataset.

The EUSTACE minus reanalysis ensemble differences are larger than the differences found between EUSTACE and in situ marine air temperatures, a global mean difference of 0.3 to 0.4 K with EUSTACE warmer than the in situ data (see the validation analysis reported in Veal et al., 2016). The EUSTACE marine test data are derived from and have been validated against nighttime data only whereas the Tmean calculated from the reanalysis data includes information about the diurnal cycle. It is possible that the EUSTACE data are overestimating the diurnal temperature range (or that the reanalyses are failing to capture the daily

maximum temperature) particularly in the cold season. The validation work showed similar patterns of larger differences at the extremes of the spatial coverage and also negative EUSTACE minus in situ differences over the western boundary currents. However, we would not expect to see larger positive differences between the EUSTACE test data and the in situ data than between the test data and the reanalyses because the reanalyses are estimates of air temperature at 2m and the in situ data are estimates of air temperature at 10m. This requires further investigation before the EUSTACE data are finalised.

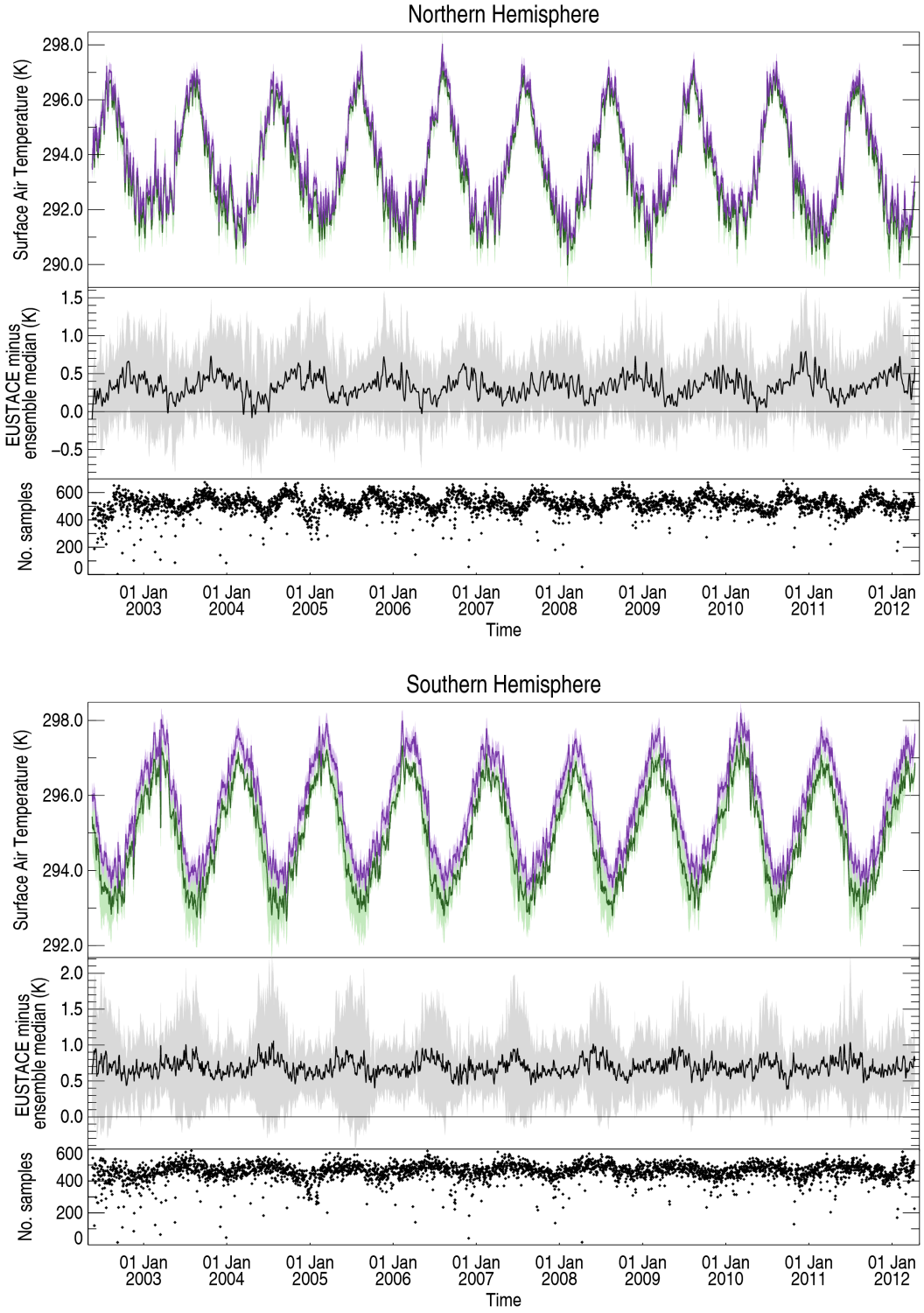


Figure 3-3: Time series of daily northern (top) and southern (bottom) hemisphere mean air temperature (K) over ocean, 2002-2012. In each panel, the reanalysis ensemble median (green line) and EUSTACE mean air temperature are plotted at the top. The middle panel shows the EUSTACE minus reanalysis ensemble median daily difference and the bottom plot shows the number of coincident 2.5° gridboxes that contribute to the hemispheric mean. In the top panel, the green shaded area is the reanalysis ensemble median +/- twice its r.s.d. and the purple shaded area is the EUSTACE mean +/- the EUSTACE 2σ uncertainty. In the middle panel, the grey shaded region is +/- twice the combined reanalysis ensemble r.s.d and EUSTACE uncertainty.

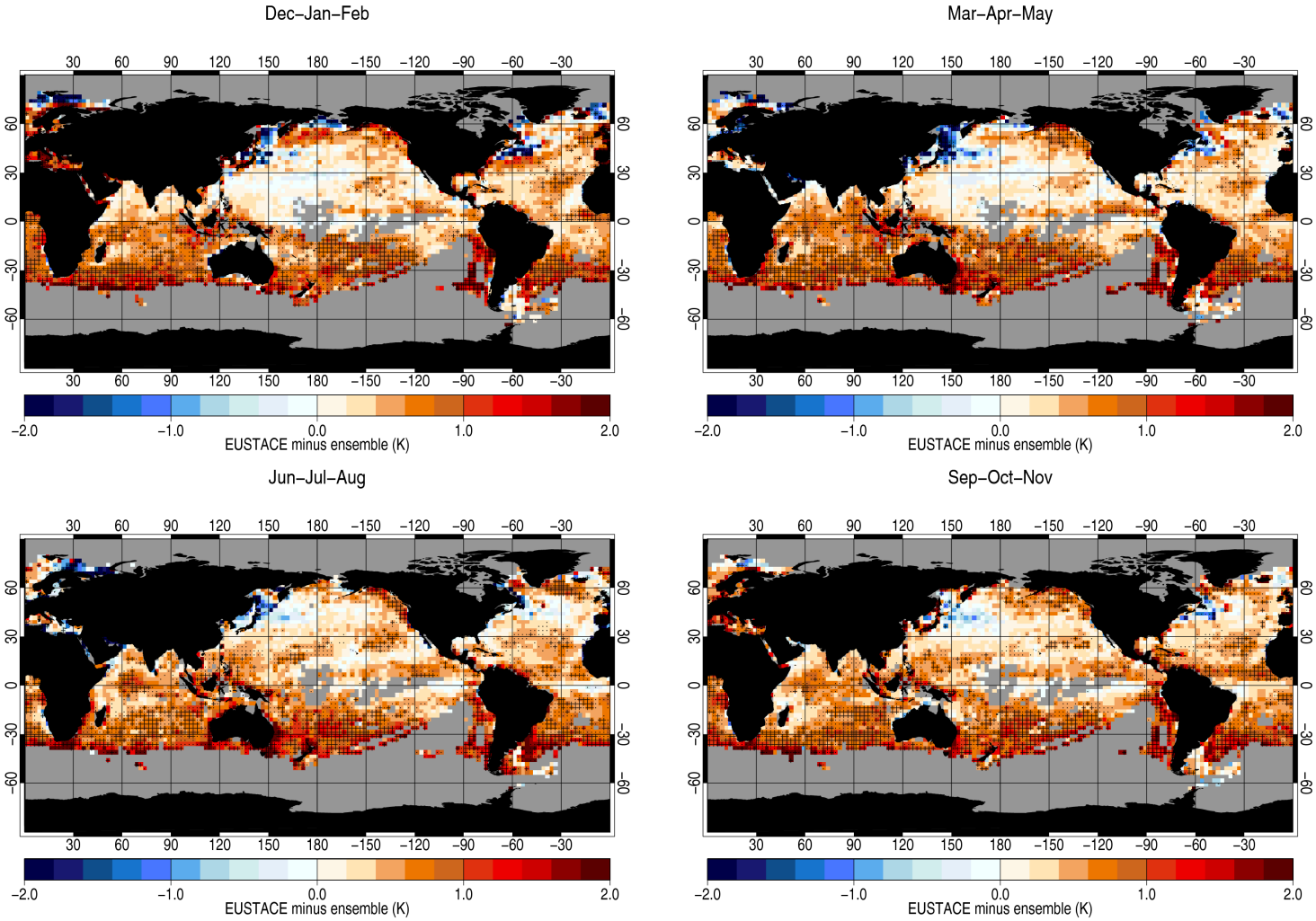


Figure 3-4: Seasonal mean over 2002-2012 of daily difference in surface air temperature over the oceans (K): EUSTACE test data minus reanalysis ensemble median . For explanation of symbols see Figure 3-2.

3.3. Ice

The EUSTACE test dataset estimates of air temperature over land ice, high concentration sea ice (over 85%), and the marginal ice zone (ice concentration between 30% and 85%) have been analysed separately. The EUSTACE test dataset masks are used to determine ice-covered areas for both the EUSTACE test data and the reanalyses. Only two years of EUSTACE ice test data were available for this intercomparison so time series are necessarily shorter and a threshold of twenty available valid data days has been applied when calculating the three-month seasonal averages. In general, mean hemispheric EUSTACE minus reanalysis ensemble median differences are larger over the ice domains than over the land and ocean. However, the differences are usually within the ±2σ uncertainty limits because the combined EUSTACE and reanalysis ensemble uncertainties are larger here. The reanalyses show poorer agreement amongst themselves in the Polar Regions compared to the rest of the globe, resulting in high reanalysis ensemble r.s.d. values here. The EUSTACE uncertainty estimates are also higher over ice, due in part to the inclusion of an additional cloud mask uncertainty component, but these are still much smaller than the reanalysis ensemble r.s.d.s. Further details of the results for the high concentration sea ice, the marginal ice zone and land ice are given in the rest of this section.

3.3.1 High concentration sea ice

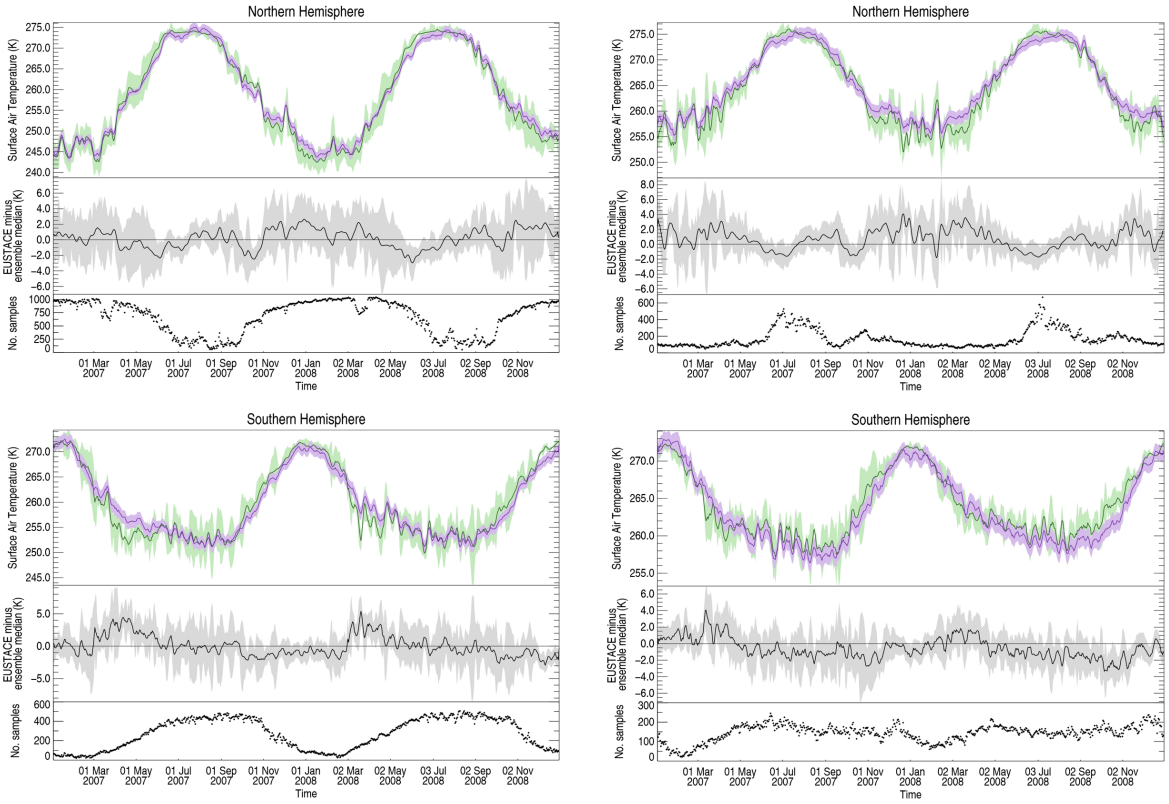
Time series of mean daily air temperature from the EUSTACE test dataset and the reanalysis ensemble median over high concentration sea ice are shown in Figure 3-5a. Sample numbers are at their lowest for high concentration sea ice in the late summer and early autumn (JAS in the northern hemisphere and JFM in the southern hemisphere) which is as expected from seasonal changes in ice extent. For nearly the whole time period, the EUSTACE minus reanalysis ensemble median difference is within the 2 sigma uncertainty estimate. The maps of seasonal mean difference (Figure 3-6) show that the highest differences occur in the region of the Canadian Archipelago and at the edge of the Greenland ice shelf. Both regions include edge lands (between land and sea in the Canadian Archipelago and between sea ice and the marginal ice zone). Sampling differences at the sub 2.5° grid box level are, therefore, likely to contribute to the discrepancy. Also, the reanalysis air temperature will be highly dependent on the specification of the sea ice in the reanalysis so, as there are large differences between the

driving sea ice datasets, we should expect differences both between individual reanalyses and between the reanalysis ensemble and the EUSTACE test dataset in the edge lands. Indeed, average differences from the ensemble median of the reanalyses over high concentration sea ice areas in the northern hemisphere appear larger than differences seen between the EUSTACE test dataset and the reference in situ data (here median discrepancy in Tmean was 0.61K, Veal et al., 2016).

In the southern hemisphere (Figure 3-7), the EUSTACE test dataset is warmer than the reanalysis ensemble median in the region of the Ross Ice Shelf and Ross sea, during March to November. Validation showed good agreement with independent drifting buoy measurements close to this region and no evidence of a persistent bias of this size (Veal et al., 2016). In the region of the Ronne Ice Shelf/Weddell Sea the EUSTACE test dataset is colder in the summer (DJF).

3.3.2 Marginal Ice Zone (MIZ)

In the MIZ, the time series (Figure 3-5b) again show that the EUSTACE test data are in reasonably good agreement with the reanalysis ensemble median and for nearly all the time series the difference between the two is within the +/- 2 sigma uncertainty range. The seasonal variation of these differences is consistent with that over high concentration sea ice regions. As is the case for the high sea ice concentration region, the MIZ shows the largest differences occurring at coastlines and at the edge of the ice zone (Figure 3-8 and Figure 3-9). Differences in the specification of the ice zone is the most likely cause although differences in sub-grid box sampling differences will also contribute.



(a) Sea-ice (b) Marginal ice zone

Figure 3-5: Time series of mean air temperature (K), 2007-2008 over (a) regions with sea-ice concentration over 85% and (b) marginal ice zones (sea ice concentration between 30% and 85%) for northern (top) and southern (bottom) hemispheres (regions include only those latitudes poleward of 60°). In each panel, the reanalysis ensemble median (green line) and EUSTACE regional mean air temperature are plotted at the top. The middle panel shows the EUSTACE minus reanalysis ensemble median daily difference and the bottom plot shows the number of coincident 2.5° gridboxes that contribute to the regional mean. In the top panel, the green shaded area is the reanalysis ensemble median +/- twice the r.s.d. and the purple shaded area is the EUSTACE mean +/- the EUSTACE 2σ uncertainty. In the middle panel, the grey shaded region is +/- twice the combined reanalysis ensemble r.s.d and EUSTACE uncertainty.

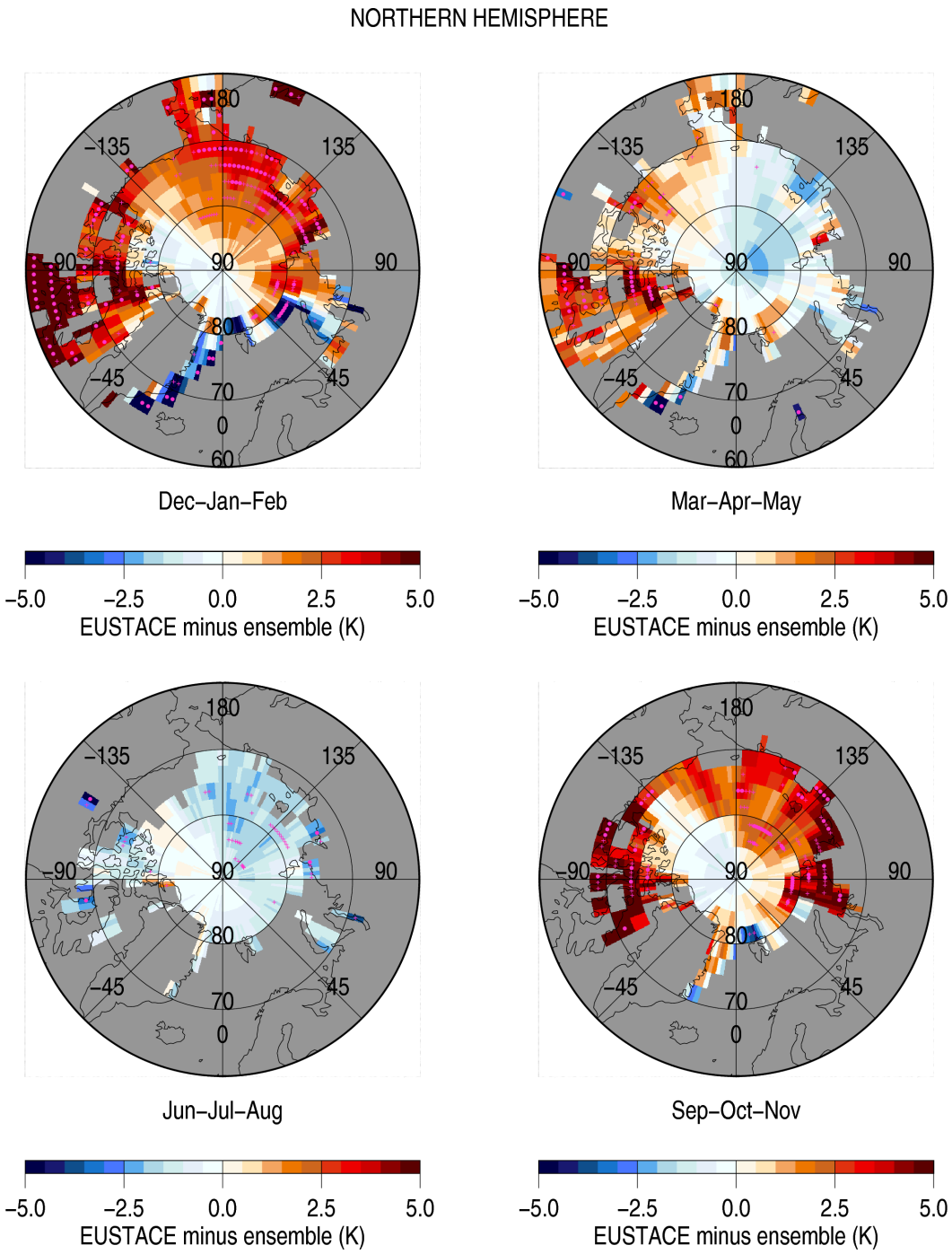


Figure 3-6: Seasonal mean daily EUSTACE test dataset minus reanalysis ensemble median difference (K) over high concentration (> 85%) sea ice in the northern hemisphere, over the period 2007-2008. Where differences exceed two (three) times the combined EUSTACE and reanalysis ensemble r.s.d the 2.5° box is marked with a cross (filled circle). Season months are indicated below discs.

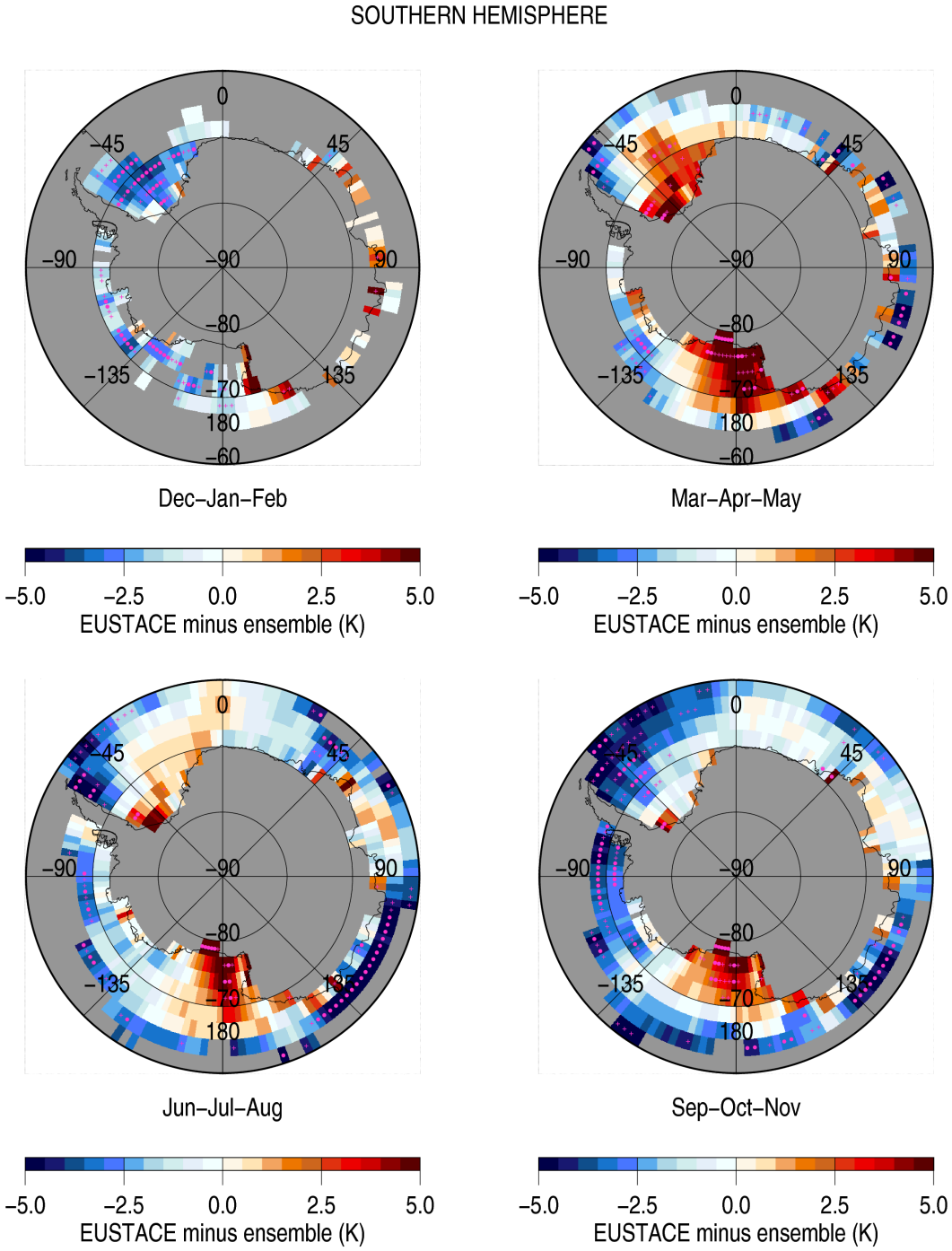


Figure 3-7: Seasonal mean daily EUSTACE test dataset minus reanalysis ensemble median difference (K) over high concentration (> 85%) sea ice in the southern hemisphere, over the period 2007-2008. For explanation of plot symbols see Figure 3-6.

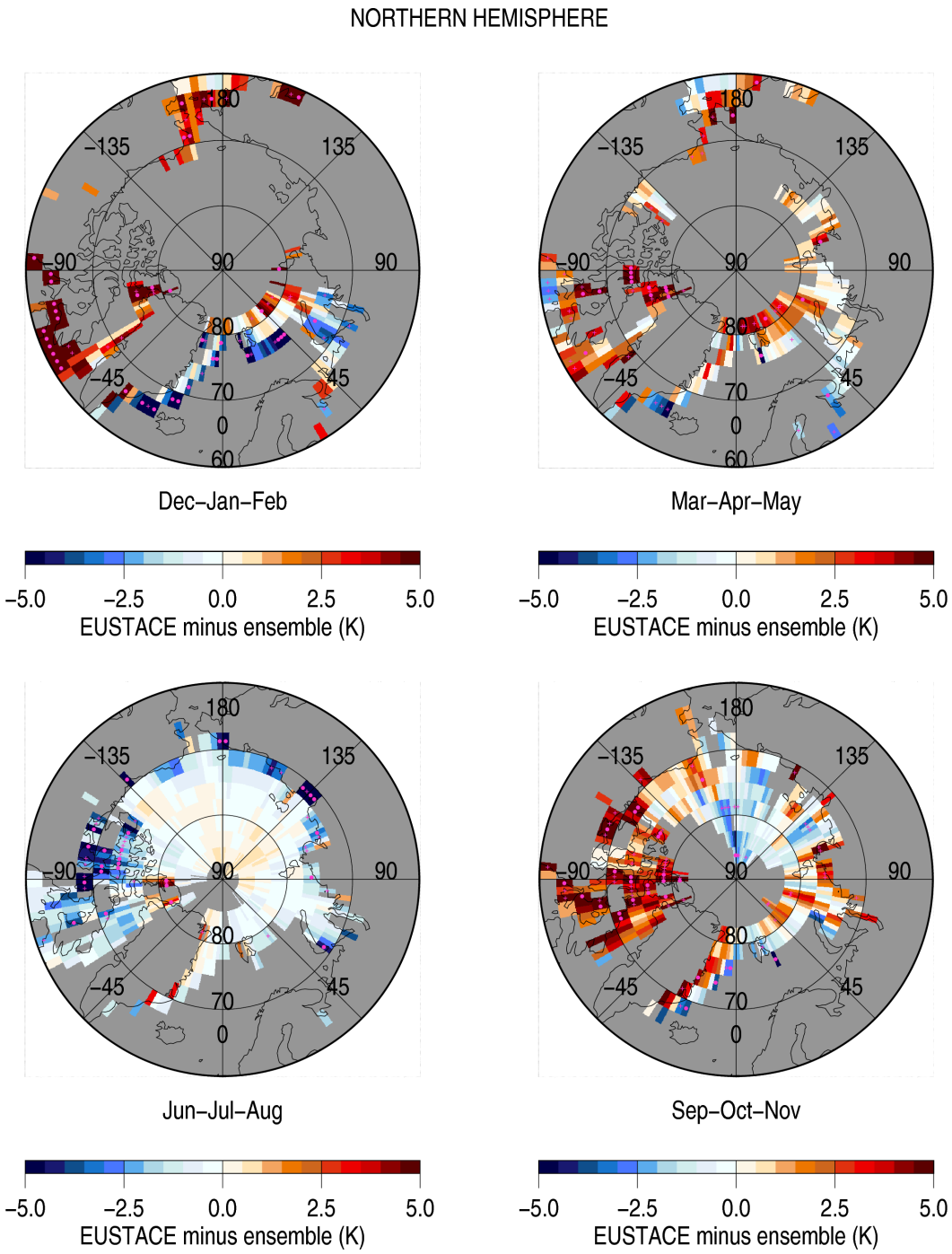


Figure 3-8: Seasonal mean daily EUSTACE test dataset minus reanalysis ensemble median difference (K) over marginal (30% to 85%) sea ice in the northern hemisphere, over the period 2007-2008. For explanation of plot symbols see Figure 3-6.

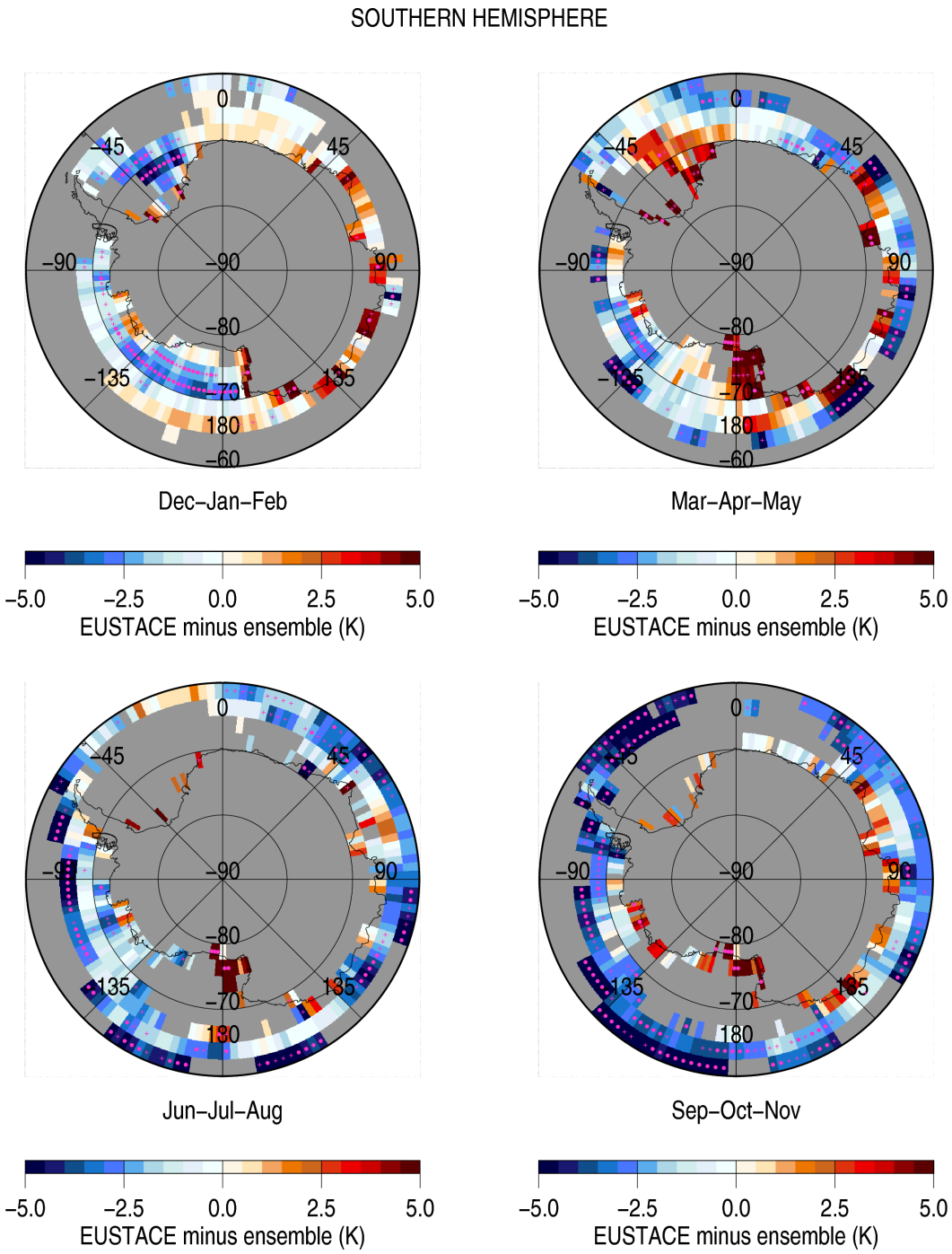


Figure 3-9: Seasonal mean daily EUSTACE test dataset minus reanalysis ensemble median difference (K) over marginal (30 to 85%) sea ice in the southern hemisphere, over the period 2007-2008. For explanation of plot symbols see Figure 3-6.

3.3.3 Land ice

Time series of mean surface air temperature estimates over Greenland and Antarctica from the EUSTACE test data and the reanalysis ensemble median are shown in Figure 3-10. In the case of Greenland, the mean EUSTACE minus reanalysis ensemble difference varies seasonally between approximately -5 K in winter and 0 to -2 K in summer. For most months of the time series, the mean EUSTACE minus reanalysis ensemble difference is within the combined 2 sigma uncertainty limits however these are large, up to 7 K. Maps of seasonal mean difference over Greenland (Figure 3-11) show that best agreement occurs in the summer (JJA) with the EUSTACE test dataset 3 to 7 K colder away from the eastern, northern, and western coasts, in other seasons. Validation against in situ data (see Veal et al., 2016) shows the EUSTACE test dataset to have a small warm bias of less than 1 K over Greenland which suggests that the reanalysis ensemble median has a larger warm bias in this region. Reanalyses have been shown to be biased warm in the cold-season, in the Arctic (Serreze et al., 2012) and this would concord with the seasonal pattern of differences between the EUSTACE test data and the reanalysis ensemble median.

Over Antarctica, sampling numbers are lowest in the winter (JJA) when cloud occurrence is highest (Figure 3-10). The EUSTACE test data show closest agreement with the reanalysis ensemble in the winter; differences are between -1 K and +1 K. However, the EUSTACE test data are colder than the reanalysis ensemble during the spring and autumn. In Figure 3-12, the EUSTACE test dataset is shown to be colder than the reanalysis ensemble median over most of East Antarctica, during the austral autumn/winter/spring. This concords with Jones and Lister (2015) who found that ERA Interim has warm biases of up to 5 K compared to inland stations on the Antarctic continent. Also, in previous validation analysis (summary results reported in Veal et al., 2016) comparison of the EUSTACE test data with in situ data from four stations in East Antarctica (Dome C, Nico, JASE7 and Panda South) found median discrepancies (EUSTACE minus in situ) between -0.3 K and 2.3 K with r.s.d.s between 2.2 K and 3.0 K, showing the EUSTACE test estimates to be more accurate than the reanalyses in this region.

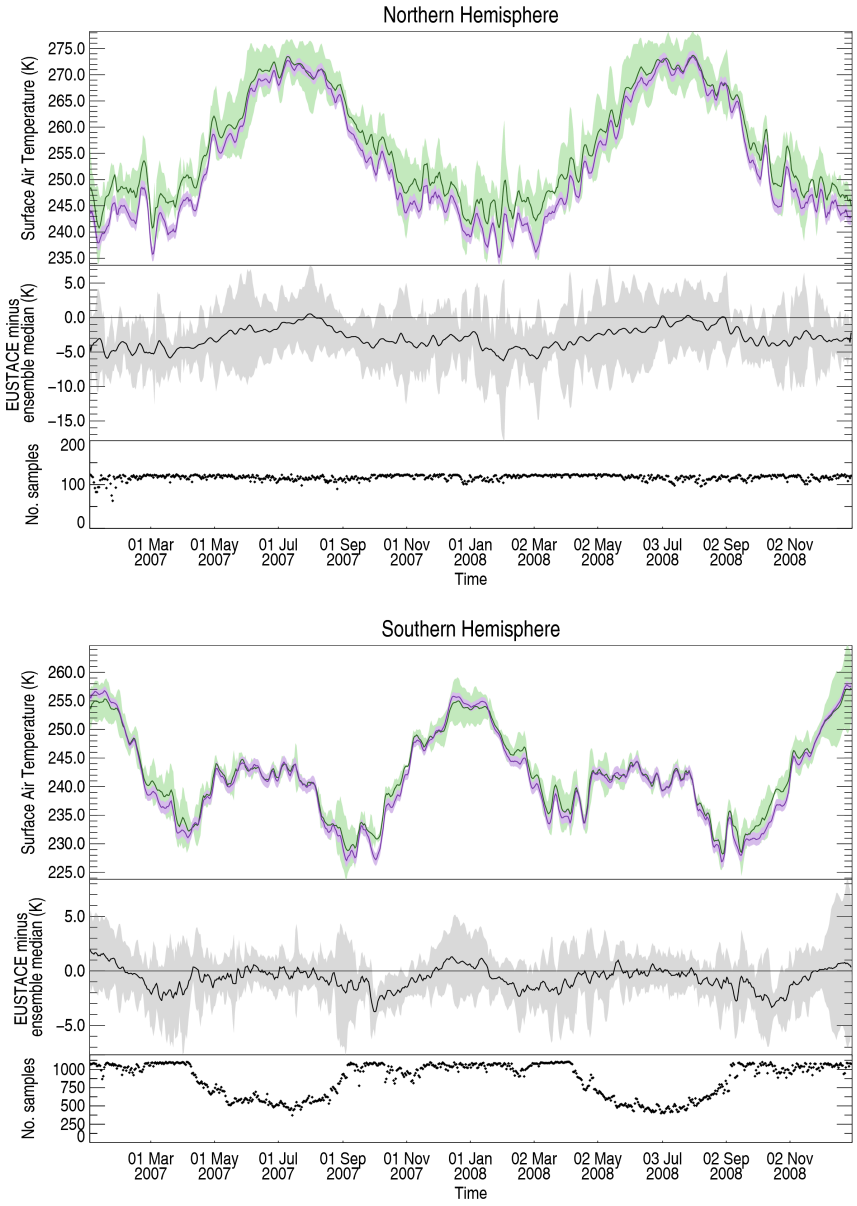


Figure 3-10: Time series of mean daily air temperature (K) over Greenland and Antarctica, 2007-2008. In each panel, the reanalysis ensemble median (green line) and EUSTACE test dataset mean regional air temperature are plotted at the top. The middle panel shows the EUSTACE minus reanalysis ensemble median daily difference and the bottom plot shows the number of coincident 2.5° gridboxes that contribute to the regional mean. In the top panel, the green shaded area is the reanalysis ensemble median +/- twice the r.s.d. and the purple shaded area is the EUSTACE mean +/- the EUSTACE 2σ uncertainty. In the middle panel, the grey shaded region is +/- twice the combined reanalysis ensemble r.s.d and EUSTACE uncertainty.See Figure 3-5 for legend

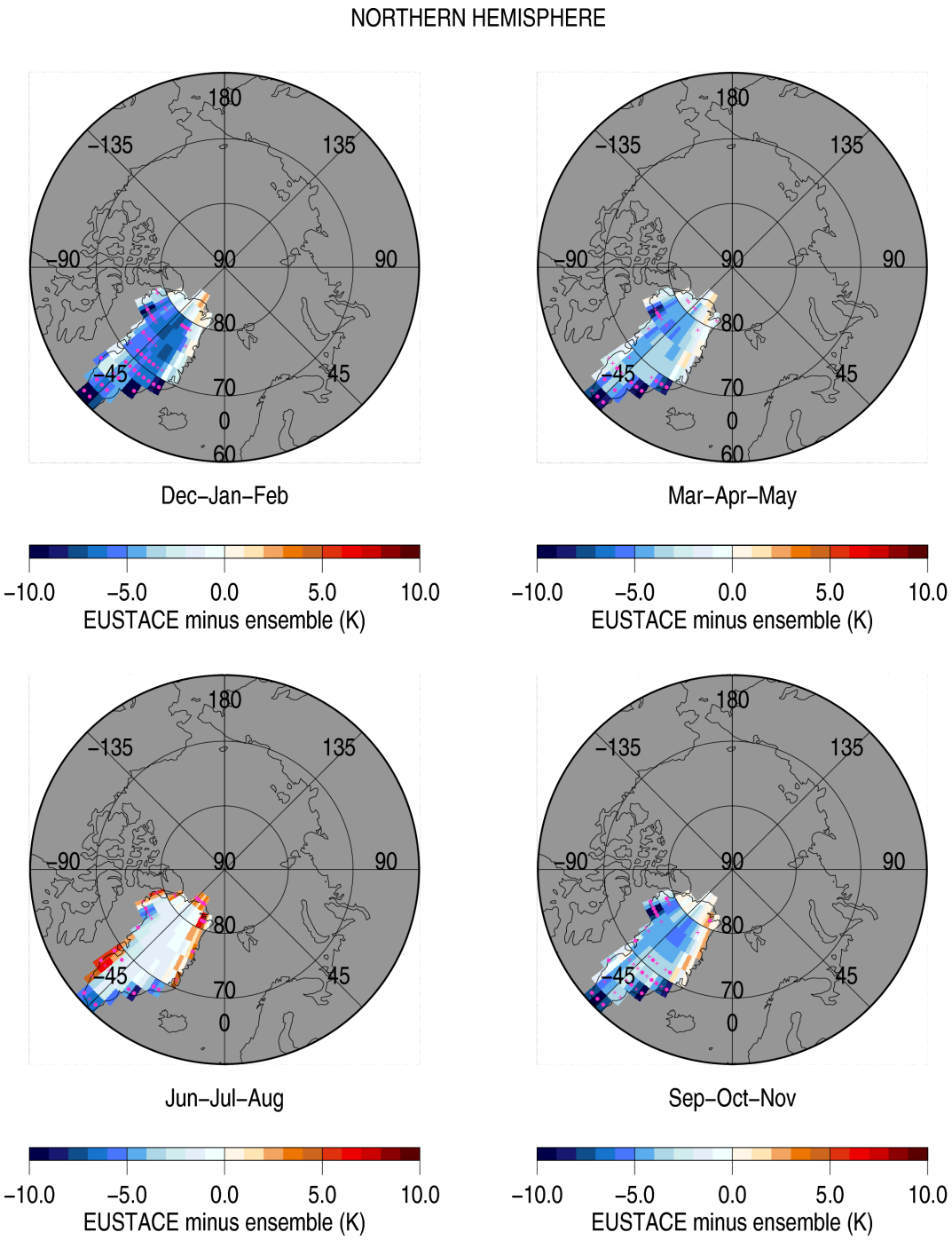


Figure 3-11: Seasonal mean daily EUSTACE test dataset minus reanalysis ensemble median difference (K) over land ice in the northern hemisphere. For explanation of plot symbols see Figure 3-6

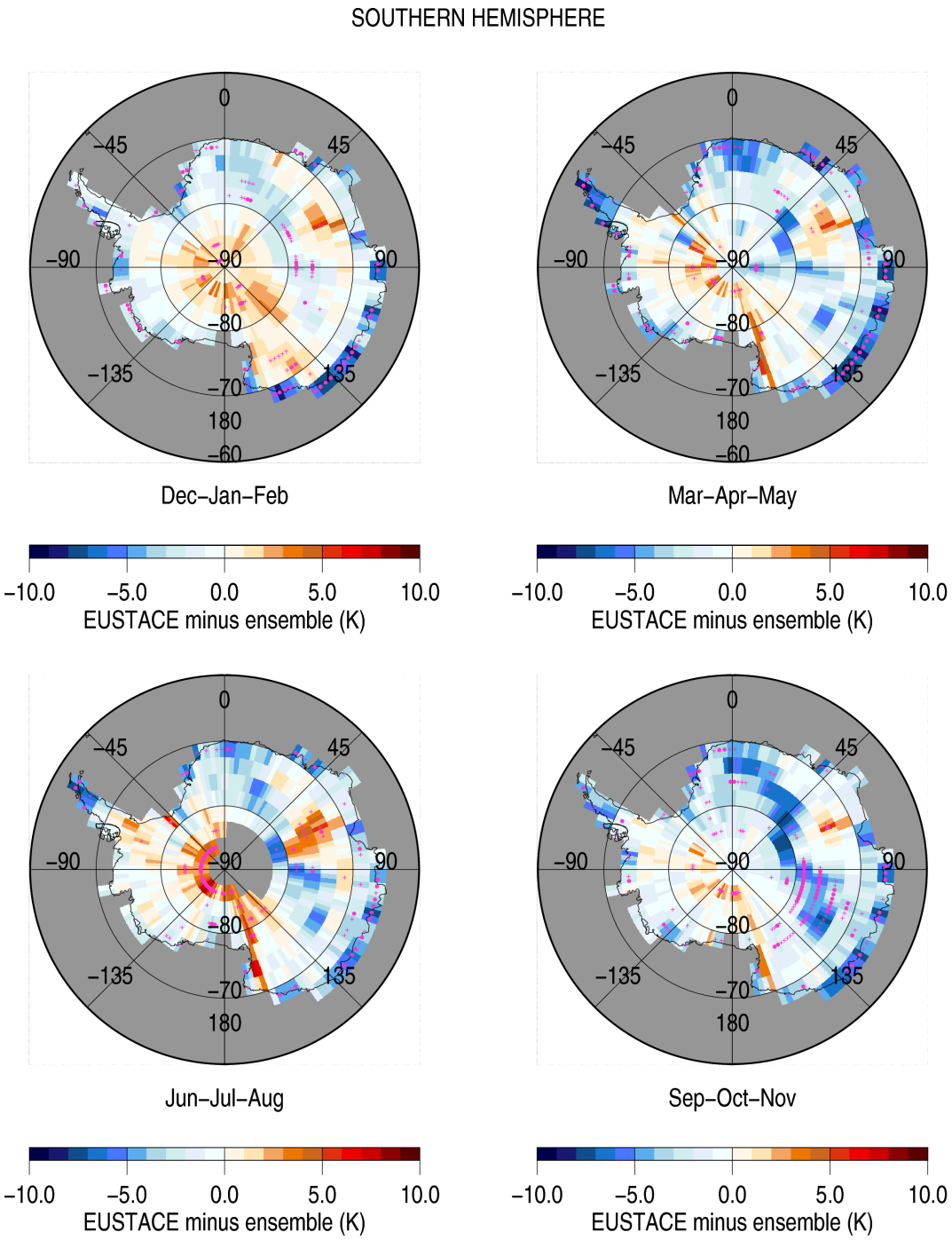


Figure 3-12: Seasonal mean daily EUSTACE test dataset minus reanalysis ensemble median difference (K) over land ice in the northern hemisphere. For explanation of plot symbols see Figure 3-6

- The EUSTACE Test Dataset estimates of daily mean surface air temperature estimated from satellite retrievals have been compared to surface air temperature from an ensemble of reanalyses on a common daily grid at a spatial resolution of 2.5° x 2.5° latitude-longitude.

- Land

    - There are noticeable differences between the EUSTACE test data and the reanalyses over land. In the northern hemisphere, the EUSTACE test data are on average colder than the reanalysis ensemble median with differences in the northern hemisphere mean exceeding 2σ in the summer.

    - EUSTACE minus reanalysis ensemble differences are often large and negative in mountainous regions. Validation against independent data in such regions in the USA showed negative differences only at the highest altitudes with positive differences at lower altitudes in these regions. Large differences are also seen in semi-arid regions and northern Asia.

    - Sampling is poor above 60°N in the winter and over the tropical forests during the wet season. Also, there are few EUSTACE test data available over Antarctica even though the input satellite LST dataset has data present.

- Ocean

    - The EUSTACE test ocean dataset is generally warmer then the reanalysis ensemble median with differences in the southern hemisphere mean exceeding 2σ for most of the time series. This result, together with the validation results indicate further investigation is needed.

    - In the region of the western boundary currents, the EUSTACE test data are colder than the reanalysis ensemble median

- Ice

    - Hemispheric mean temperatures from the EUSTACE test ice dataset and the reanalysis ensemble median are in agreement within 2σ uncertainty limits over land ice, high concentration sea ice and marginal zones. The

uncertainties on both the EUSTACE and the reanalysis ensemble hemispheric means tends to be larger over the ice domains compared with the land and ocean domains.

- Over Greenland and inland Antarctica EUSTACE temperatures are colder than the reanalysis ensemble in certain seasons. Previous studies which have shown the reanalyses to have large warm biases in these regions.

- The EUSTACE test dataset is warmer than the reanalysis ensemble median over the sea ice of the Ross Ice Shelf and Ross Sea during March to November. There is no evidence from comparison to independent reference data that there is any large bias in the EUSTACE test dataset here. In the region of the Ronne Ice Shelf/Weddell Sea the EUSTACE test dataset is colder in the austral summer.

4.Data sources

European Centre for Medium-Range Weather Forecasts' ERA-Interim Project. European Centre for Medium-Range Weather Forecasts. (Updated monthly). Accessed 21 September 2016. **http://apps.ecmwf.int/datasets/**

European Centre for Medium-Range Weather Forecasts (2104). European Centre for Medium-Range Weather Forecasts' atmospheric reanalysis of the 20th century (ERA-20C). European Centre for Medium-Range Weather Forecast. Accessed 21 September 2016. **http://apps.ecmwf.int/datasets/**

Japan Meteorological Agency/Japan (2013), JRA-55: Japanese 55-year Reanalysis, Daily 3-Hourly and 6-Hourly Data, http://dx.doi.org/10.5065/D6HH6H41, Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, Colo. (Updated monthly.) Accessed 03 November 2016. **http://rda.ucar.edu/datasets/ds628.0/**

Global Modeling and Assimilation Office (GMAO), Modern-Era Retrospective Analysis For Research And Applications (MERRA). NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC). Accessed 05 December 2015. **http://disc.sci.gsfc.nasa.gov**

National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce (1994), NCEP/NCAR Global Reanalysis Products, 1948-continuing, http://rda.ucar.edu/datasets/ds090.0/, Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, Colo. (Updated monthly.) Accessed 07 July 2016. **http://www.esrl.noaa.gov/**

5.References

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, 137(656), 553-597, doi: 10.1002/qj.828.

Huber, P. J. and Ronchetti, E. M. (2009), Robust Statisitics. Wiley, NewYork.

Jones, P. D., and D. H. Lister (2015), Antarctic near-surface air temperatures compared with ERA-Interim values since 1979, International Journal of Climatology, 35(7), 1354-1366, doi: 10.1002/joc.4061.

Kalnay, E., et al. (1996), The NCEP/NCAR 40-Year Reanalysis Project, Bulletin of the American Meteorological Society, 77(3), 437-471, doi: doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kobayashi, S., Y. Ota, Y. Harada, A. Ebita, M. Moriya, H. Onoda, K. Onogi, H. Kamahori, C. Kobayashi, H. Endo, K. Miyaoka, and K. Takahashi, 2015: The JRA-55 Reanalysis: General Specifications and Basic Characteristics. J. Met. Soc. Jap., 93(1), 5-48 (DOI: 10.2151/jmsj.2015-001).

Poli, P., et al. (2013), The data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C), in ERA Report Series, edited, p. 59, ECMWF, Shinfield Park, Reading.

Rienecker, M. M., et al. (2011), MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, Journal of Climate, 24(14), 3624-3648, doi: 10.1175/JCLI-D-11-00015.1.

Serreze, M. C., A. P. Barrett, and J. Stroeve (2012), Recent changes in tropospheric water vapor over the Arctic as assessed from radiosondes and atmospheric reanalyses, J. Geophys. Res., 117(D10), D10104, doi: 10.1029/2011jd017421.

Veal, K and G. Corlett (2016), Validation report for the intermediate fields inferred from retrievals, EUSTACE Deliverable 3.3.